Error in Monte Carlo, quasi-error in Quasi-Monte Carlo

Ronald Kleiss[1] **and Achilleas Lazopoulos**[2]

IMAPP
Institute of Mathematics, Astrophysics and Particle Physics
Radboud University, Nijmegen, the Netherlands

Abstract

While the Quasi-Monte Carlo method of numerical integration achieves smaller integration error than standard Monte Carlo, its use in particle physics phenomenology has been hindered by the abscence of a reliable way to estimate that error. The standard Monte Carlo error estimator relies on the assumption that the points are generated independently of each other and, therefore, fails to account for the error improvement advertised by the Quasi-Monte Carlo method. We advocate the construction of an estimator of stochastic nature, based on the ensemble of pointsets with a particular discrepancy value. We investigate the consequences of this choice and give some first empirical results on the suggested estimators.

1 Monte Carlo and Quasi-Monte Carlo

1.1 Introduction

In numerical integration, the main problem is not to obtain a numerical answer[3] for the integral, but rather, on the one hand, to ensure that the inherent numerical error is as small as possible, and, on the other hand, to estimate this error as precisely as possible. For integrands with well-known smoothness properties, *a-priori* estimates of the numerical error are possible, but for most practical applications the smoothness properties of the integrand can only be investigated in the course of the integration itself, that is, by repeated numerical evaluation of the integrand.

In this paper, we shall be concerned with the integration errors arising in Monte Carlo and Quasi-Monte Carlo integration. In these methods, the integration

[1] R.Kleiss@science.ru.nl
[2] A.Lazopoulos@science.ru.nl
[3] which is known to be 42, see [1].

nodes are distributed in a (more or less) stochastic manner, and the integration errors are therefore of an essentially probabilistic nature. The difference between Monte Carlo and Quasi-Monte Carlo is that in the former, the integration points are iid[4] uniform in the integration region[5], while in the latter the integration points are not chosen independently, but rather with an explicit interdependence so that their overall distribution is 'smoother', in a sense discussed below.

In stochastic integration methods of the Monte Carlo or Quasi-Monte Carlo types, the integration error is itself an estimate, which contains its own error. That this is not an academic point becomes clear when we realize that the error estimate is routinely used to provide *confidence levels* for the integral estimate (be it based either on Chebyshev or Central-Limit-Theorem, Gaussian rules); and a mis-estimate of the integration error can lead to a serious under- or overestimate of the confidence level. As an example, suppose that the Central Limit Theorem is applicable, so that the integration result is drawn from a Gaussian distribution centered around the true integral value. One standard deviation, as estimated by Monte Carlo, corresponds to a two-sided confidence level of 68%. If the error estimate is off by 50% (admittedly a large value), the actual confidence level may then be anything between 38% and 87%.

From this consideration, we are therefore led to a hierarchy of error estimates: the *first-order* error is that on the integral estimate, while the *second-order* error is the error on the error estimate. This in turn has, of course, its own *third-order* error, and so on. Higher orders than the second one, however, appear to be too academic for practical relevance, but we should like to argue that, in any serious integration problem, the second-order error ought to be included. In what follows we shall discuss the first- and second-order error estimates.

Due to the absence of a Quasi-Monte Carlo error estimator, users of Quasi--Monte Carlo have been estimating the integration error with the classical Monte Carlo formula, as if the point set was iid. This systematically overestimates the error in any case where the quasi point-set is of any worth. Moreover, no confidence levels can be assigned since the classical estimator does not average to the error made by the quasi, non-iid point-sequence. The purpose of this paper is

[4]iid stands for 'independent, identically distributed'.

[5]This ignores the possible interpretation of stratified and importance sampling methods of variance reduction. These can, at any rate, always be formulated in terms of methods using iid uniform integration points.

to investigate possible estimators for Quasi-Monte Carlo integration taking under consideration the non-iid nature of the underlying point-set[6].

1.2 Monte Carlo estimators

In this section we briefly review the probabilistic theory underlying Monte Carlo integration. This is of course well known, but we include it here so that the significant difference with Quasi-Monte Carlo can become clear.

Throughout this paper we shall consider integration problems over the d-dimensional unit hypercube $C = [0,1)^d$. The integrand is a function $f(\vec{x})$, which we shall assume real and non-negative, and, of course, integrable over C. We shall define

$$J_m = \int_C f(\vec{x})^m \, d^d\vec{x} \quad , \quad m = 1, 2, 3, \ldots \quad , \tag{1}$$

so that J_1 is the required integral. Note that J_m is not necessarily finite for $m \geq 2$. In Monte Carlo we assume N integration points, to be chosen iid from the uniform probability distribution over C. This means that the *point set* $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N\}$ on which the integration is based is *assumed* to be a typical member of an ensemble of such point sets, in such a way that the combined probability distribution of the N points over this ensemble is the uniform iid one:

$$P_N(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = 1 \quad . \tag{2}$$

We shall take the averages over this ensemble. Other assumptions on the underlying ensemble from which the point set X is believed to be chosen are possible, leading to a different form of P_N. In this, the situation is not different from that encountered in statistical mechanics. The above assumption, however, is the one that is always made in regular Monte Carlo and is justified to some extent by the fact that good-quality (pseudo)random number generators are actually available, allowing us to build ensembles of point sets X that indeed have the above property (2).

[6]The opposite direction - re-introducing randomness by reshuffling the points of the Quasi--Monte Carlo sequence in a way that preserves their uniformity properties, thus allowing for the use of a 'classical'-type estimator - has been studied extensively in the literature (see [9] and references therein). Such point-sequences behave better than Monte Carlo sequences and, for integrands with certain properties, as good as Quasi-Monte Carlo sequences. Estimating the error, however, requires the use of a number r of different reshufflings of a point-set with n points, thereby trading off accuracy for knowledge of the error.

Let us assume that a point set X has been generated, and the values of the integrand $f(\vec{x})$ at all these points have been computed. These we shall denote by $f_j \equiv f(\vec{x}_j)$, $j = 1, 2, \ldots, N$. From these we can compute the discrete analogues of the integrals J_m, which are computable in linear time (that is, time proportional to N):

$$S_m = \sum_{j=1}^{N} (f_j)^m \quad . \tag{3}$$

The Monte Carlo estimate of the integral is then

$$E_1 = \frac{1}{N} S_1 \quad . \tag{4}$$

The expected value of E_1 over the above ensemble of point sets is given by

$$\langle E_1 \rangle = \frac{1}{N} \sum_i \langle f_i \rangle = \int_C f(\vec{x}) \; d^d\vec{x} = J_1 \quad , \tag{5}$$

which is indeed the required integral: this is the basis for the Monte Carlo method. Its usefulness appears if we compute the variance of E_1:

$$\sigma(E_1)^2 = \left\langle E_1^2 \right\rangle - \langle E_1 \rangle^2 = \frac{1}{N} \left(J_2 - J_1^2 \right) \quad . \tag{6}$$

Since this decreases as N^{-1}, the Monte Carlo method actually converges for large N. Note that the leading, $\mathcal{O}(N^0)$, terms of $\langle E_1^2 \rangle$ and $\langle E_1 \rangle^2$ cancel against each other: this is a regular phenomenon in variance estimates of this kind[7]. The variance $\sigma(E_1)^2$ is estimated by the first-order error estimator (also called 'classical' or 'pseudo' estimator in what follows)

$$E_2 = \frac{1}{N^2} S_2 - \frac{1}{N^3} S_1^2 \quad , \tag{7}$$

for which we have

$$\langle E_2 \rangle = \sigma(E_1)^2 + \mathcal{O}(N^{-2}) \quad . \tag{8}$$

Since N is usually quite large, at least 10,000 or so, we feel justified in working only to leading order in N. The squared error of E_2 is computed to be, to leading order in N,

$$\sigma(E_2)^2 = \frac{1}{N^3} \left(J_4 - 4 J_3 J_1 - J_2^2 + 8 J_2 J_1^2 - 4 J_1^4 \right) \quad , \tag{9}$$

[7] It should be pointed out that what we estimate is the average of the squared error, rather than the error itself, and squaring and averaging do *not* commute. In fact, this is another reason why the second-order estimate is relevant.

for which the estimator is

$$E_4 \;=\; \frac{1}{N^7}\left(N^3 S_4 - 4N^2 S_3 S_1 - N^2 S_2^2 + 8N S_2 S_1^2 - 4S_1^4\right) \quad . \tag{10}$$

which can also be computed in linear time; we have

$$\langle E_4 \rangle = \sigma\,(E_2)^2 + \mathcal{O}(N^{-4}) \quad . \tag{11}$$

Some details on the computation of leading-order expectation values of this type, as well as (for purposes of illustration) the form of the third- and fourth-order error estimators, E_8 and E_{16}, respectively, are given in the Appendix.

A final remark is in order here. The Central Limit theorem, which ensures that the error estimate can be used to derive *Gaussian* confidence levels, can also be inferred from the computation of the higher cumulants of the error distribution: we find for the skewness

$$\left\langle (E_1 - \langle E_1 \rangle)^3 \right\rangle = \frac{1}{N^2}\left(J_3 - 3J_2 J_1 + 2J_1^3\right) \quad , \tag{12}$$

and the unnormalized kurtosis:

$$\left\langle (E_1 - \langle E_1 \rangle)^4 \right\rangle - 3\sigma\,(E_1)^2 = \frac{1}{N^3}\left(J_4 - 4J_3 J_1 - 3J_2^2 + 12 J_2 J_1^2 - 6J_1^4\right) \quad , \tag{13}$$

which indicate that the higher cumulants decrease faster than the variance with increasing N; we shall examine this later on for the case of Quasi-Monte Carlo.

1.3 Quasi-Monte Carlo estimators

1.3.1 Multi-point distribution and correlation functions

In contrast to the case of regular Monte Carlo, the technique of Quasi-Monte Carlo relies on point sets in which the points are *not* chosen iid from the uniform distribution, but rather interdependently. To make this more specific, let us consider a point set X of N points. For each such a point set, we may define a *measure of non-uniformity*, called a *discrepancy* or, as in this paper, a *diaphony*. Its precise definition is presented below: for now, suffice it to demand that there exist a function $D(X)$ of the point set, which increases with its non-uniformity: $D(X) = 0$ if the point set is perfectly uniform in all possible respects, an ideal situation that can never be obtained for any finite point set. The Quasi-Monte Carlo method

consists of using point sets X for which $D(X)$ has some value s which is (very much) smaller than $\langle s \rangle$, the value that may be expected for truly iid uniform ones.

Given that such 'quasi-random' point sets can be obtained, how does one use them in numerical integration? The obvious issue here is to determine of what ensemble the quasi-random point set X can be considered to be a 'typical' member. In this paper, we should like to advocate the viewpoint that, since the main additional property of the quasi-random point set that distinguishes it from truly random point sets is its 'anomalously small' discrepancy D, the ensemble ought to consist of those point sets that are iid uniformly, with the additional constraint that the discrepancy D has the particular value $D(X) = s$ for the actually used point set[8]. On this premise, the Quasi-Monte Carlo analogue of Eq.(2) would then be the assumption

$$P_N(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = \frac{1}{H(s)} \delta(D(X) - s) \quad , \tag{14}$$

where s is, again, the observed value of the discrepancy of X, on which P_N must now of course depend; and $H(s)$ is the probability density to happen upon a point sets X with this discrepancy in the regular-Monte Carlo ensemble:

$$H(s) = \int_C \delta(D(X) - s) \; d^d\vec{x}_1 \; d^d\vec{x}_2 \; d^d\vec{x}_N \tag{15}$$

The actual computation of $H(s)$ for given definition of the discrepancy is referred to the next section. What interests us here is the fact that P_N is now no longer simply unity, since that would imply independence of the points in the point set. Let us therefore write the *multi-pont distribution* as

$$P_N(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = 1 - \frac{1}{N} F_2(s; \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) \quad , \tag{16}$$

where we have anticipated a factor $1/N$ in the *multi-point correlation* F.

1.3.2 Properties of the correlation function

Since the value of the discrepancy of a given point-set X, should be independent of the order in which the points are generated, $F_k(s; \vec{x}_1 \ldots \vec{x}_k)$ must be totally

[8] We do not examine the possible alternative that the point sets in the ensemble must have discrepancy *in the neighborhood* of the observed value s; this amounts to the distinction between the micro-canonical and the canonical ensemble in statistical mechanics.

symmetric; moreover, we must have

$$F_k(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_K) = \int_C F_{k+1}(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_k,\vec{x}_{k+1})\ d^d\vec{x}_{k+1}\ \ , \tag{17}$$

which is not as trivial as it might seem since the value of the discrepancy, s, is based on the full N points and not on the smaller set of k or $k+1$ points. Finally, for the Quasi-Monte Carlo integral to be unbiased, we must have

$$P_1(s;\vec{x}_1) = 1\ \ , \tag{18}$$

so that

$$\int_C F_2(s;\vec{x}_1,\vec{x}_2)\ d^d\vec{x}_2 = 0\ \ . \tag{19}$$

These remain, of course, to be proven and we shall do so in the next section, for a particular choice of discrepancy. Moreover, we shall show there that the multi-point correlation F_N is, to leading order in $1/N$, made up from two-point correlations F_2:

$$F_k(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_k) = \sum_{1\le m<n\le k} F_2(s;\vec{x}_m,\vec{x}_n)\ \ . \tag{20}$$

This establishes the properties of our ensemble of point sets X on which, in our view, the Quasi-Monte Carlo estimates ought to be based.

1.3.3 Estimators

We shall indicate the 'Quasi-Monte Carlo' nature of the estimators by the superscript (q). The first estimator is that of the integral:

$$E_1^{(q)} = \frac{1}{N}\sum f_j\ \ . \tag{21}$$

Here, and in the rest of this section, the sums will run from 1 to N. Denoting by the subscript (q) averages with respect to the 'quasi-random' ensemble discussed above, we then have

$$\left\langle E_1^{(q)}\right\rangle_{(q)} = \int_C f(\vec{x})\ P_1(s;\vec{x})\ d^d\vec{x} = J_1\ \ , \tag{22}$$

as before: owing to the fact that the one-point distribution is uniform, the Quasi-Monte Carlo estimate is indeed as unbiased as the Monte Carlo one. The distinction between the two methods appears in the first-order error estimate. Let us define

$$\alpha(\vec{x}_i, \vec{x}_j) = 1 + F_2(s; \vec{x}_i, \vec{x}_j) \ ; \tag{23}$$

then, we have

$$\sigma\left(E_1^{(q)}\right)^2_{(q)} = \frac{1}{N}\left(J_2 - \int f_1 f_2 \alpha_{12}\right) \quad + \quad \mathcal{O}\left(\frac{1}{N^2}\right) \ . \tag{24}$$

where we have adopted the straightforward convention for integrals

$$\int f_1 f_2 \alpha_{12} = \int_C f(\vec{x}_1)\ f(\vec{x}_2)\ \alpha(\vec{x}_1, \vec{x}_2)\ d^d\vec{x}_1\ d^d\vec{x}_2 \ , \tag{25}$$

etcetera. As before, we shall insouciantly neglect terms that are sub-leading in $1/N$. The advantage of the Quasi-Monte Carlo method is now clear: if we can ensure that $\alpha_{12} > 1$ 'where it counts', that is, generally, when $\vec{x}_1$ and $\vec{x}_2$ are 'close' in some sense, then the Quasi-Monte Carlo error will be smaller than the Monte Carlo one. A good Quasi-Monte Carlo point set, therefore, is one in which the points 'repel' each other to some extent.

The first-order error estimate is now simply

$$E_2^{(q)} = \frac{1}{N^2}\sum f_i^2 - \frac{1}{N^3}\sum f_i f_j \alpha_{ij} \ . \tag{26}$$

It is simple to show that, indeed

$$\left\langle E_2^{(q)} \right\rangle_{(q)} = \sigma\left(E_1^{(q)}\right)^2_{(q)} + \mathcal{O}(N^{-2}) \ ; \tag{27}$$

however, *evaluating* $E_2^{(q)}$ is less trivial since it is not obvious how to do this in time linear in N. We shall discuss this later. The variance of the estimator $E_2^{(q)}$ can be evaluated to

$$\begin{aligned} \sigma\left(E_2^{(q)}\right)^2 \ = \ & \frac{1}{N^3}\left(\int f_i^4 - 4\int f_i^3 f_j \alpha_{ij} - \int f_i^2 f_j^2 \alpha_{ij}\right. \\ & +4\int f_i^2 f_k f_l \alpha_{ik}\alpha_{kl} + 4\int f_i^2 f_k f_l \alpha_{ik}\alpha_{il} \\ & \left. -4\int f_i f_j f_k f_l \alpha_{ij}\alpha_{jk}\alpha_{kl}\right) + \mathcal{O}(N^{-4}) \ , \end{aligned} \tag{28}$$

for which the corresponding estimator (to leading order) is

$$\begin{aligned} E_4^{(q)} &= \frac{1}{N^7}\Big(N^3\sum f_i^4 - 4N^2\sum f_i^3 f_j \alpha_{ij} - N^2\sum f_i^2 f_j^2 \alpha_{ij} \\ &\quad +4N\sum f_i^2 f_k f_l \alpha_{ik}\alpha_{kl} + 4N\sum f_i^2 f_k f_l \alpha_{ik}\alpha_{il} \\ &\quad -4\sum f_i f_j f_k f_l \alpha_{ij}\alpha_{jk}\alpha_{kl}\Big) \quad . \end{aligned} \tag{29}$$

The details of this computation are discussed in the Appendix. It goes without saying that the substitution $\alpha_{ij} \to 1$ will reduce all the Quasi-Monte Carlo results to the regular Monte Carlo ones.

We can now see why the 'classical' estimator Eq.(7) overestimates the error. Under the quasi distribution P_2 of Eq.(16) the classical estimator averages to

$$\begin{aligned} \langle E_2\rangle_{(q)} &= \left\langle \frac{1}{N^2}\sum_i f_i^2 - \frac{1}{N^3}\sum_{i,j} f_i f_j \right\rangle_{(q)} \\ &= \frac{1}{N}(J_2 - J_1^2) - \frac{1}{N^2}\int f(x)f(y)F(x,y) + \mathcal{O}(\frac{1}{N^2}) \end{aligned} \tag{30}$$

The term involving the correlator is suppressed by $\frac{1}{N}$, which shows that E_2 averages to something different than the variance of E_1 under the quasi distribution. Moreover, we will show in section 2.3 that[9] the integral of the suppressed term is strictly positive for any point-set that is better than a truly random one. So E_2 omits a strictly negative term when estimating the error.

While it is true that the estimator Eq.(26) averages to a quantity whose leading order in N is equal to the leading order of $\sigma\left(E_2^{(q)}\right)^2$, it suffers from the following disagreeable property: for a constant integrand, while the first two terms vanish identically, the third approaches zero asymptotically from negative values. This leads to a negative squared error for all practical purposes. Although this is not disastrous per se, it indicates the reason for the appearance of negative errors also for non-constant integrands, as will become apparent once we have a concrete expression for the correlation function. It is, thus, desirable to obtain an estimator that vanishes identically for constant functions. This is achieved by

$$E_2^{(q_2)} = \frac{1}{N^{\underline{2}}}\sum_i f_i^2 - \frac{1}{NN^{\underline{2}}}\sum_{i,j}^{\wedge} f_i f_j - \frac{1}{NN^{\underline{4}}}\sum_{i,j,k,l}^{\wedge} f_i f_j (F_{i,j} - F_{i,k} - F_{l,j} + F_{l,k}) \tag{31}$$

[9] under fairly general conditions for the function $f(x)$.

where the $\hat{\Sigma}_{i,j...}$ denotes a sum with all indices different, and $F_{i,j} \equiv F_2(s;\vec{x}_i,\vec{x}_j)$. This quantity averages to

$$\left\langle E_2^{(q_2)} \right\rangle = \frac{1}{N}(J_2 - J_1^2) - \frac{1}{N}\int dxdydzdw\ f(x)f(y)\left[F_{x,y} - F_{x,w} - F_{z,y} + F_{z,w}\right] \tag{32}$$

which equals the leading part of $\sigma\left(E_2^{(q)}\right)^2$ thanks to Eq.(19). It is easy to check that the estimator of Eq.(31) vanishes identically for a constant integrand and any N, thanks to the antisymmetry property of the quadruple sum.

1.3.4 Cumulants of E_1

As a final remark, we may also investigate the cumulants of the Quasi-Monte Carlo estimator E_1. We write the expansion of the correlation function F_k over $1/N$ as

$$F_k(s;\vec{x}_1,\ldots,\vec{x}_k) \equiv F_k^{(1)} + \frac{1}{N}F_k^{(2)} + \frac{1}{N^2}F_k^{(3)} + \ldots \tag{33}$$

and define

$$\mathcal{M}^{(a)}_{i_1,\ldots,i_k} \equiv \int f(\vec{x}_1)^{i_1}\ldots f(\vec{x}_k)^{i_k}F_k^{(a)}(s;\vec{x}_1,\ldots,\vec{x}_k) \tag{34}$$

It is evident that if Eq.(20) holds, we have

$$\mathcal{M}^{(1)}_{1,1,\ldots,1} = \frac{k!}{2}J_1^{k-2}\mathcal{M}^{(1)}_{1,1} \tag{35}$$

The cumulants are defined as

$$c_n = \left\langle \left(E_1^{(q)} - \left\langle E_1^{(q)} \right\rangle_{(q)} \right)^n \right\rangle_{(q)} \tag{36}$$

The variance of E_1 is then

$$c_2 = \frac{1}{N}(J_2 - J_1^2 - \mathcal{M}^{(1)}_{1,1}) + O(\frac{1}{N^2}) \tag{37}$$

The skewness is

$$c_3 = \frac{1}{N^2}(J_3 - 3J_1J_2 + 2J_1^3 - 3\mathcal{M}^{(1)}_{1,2} + 3J_1\mathcal{M}^{(2)}_{1,1} + 6J_1\mathcal{M}^{(1)}_{1,1} - \mathcal{M}^{(2)}_{1,1,1}) + O(\frac{1}{N^3}) \tag{38}$$

The unnormalized kurtosis is

$$c_4 - 3c_2^2 = \frac{1}{N^2}(-\mathcal{M}^{(2)}_{1,1,1,1} - 3(\mathcal{M}^{(1)}_{1,1})^2 + 4J_1\mathcal{M}^{(2)}_{1,1,1} - 6J_1^2\mathcal{M}^{(2)}_{1,1}) + O(\frac{1}{N^3}) \tag{39}$$

The above results indicate that a correlation function that satisfies the property of Eq.(20) leads to a distribution whose skewness decreases faster with N than does the variance, but when it comes to the kurtosis (and higher cumulants), additional properties regarding the next-to-leading order expression for F (denoted above by $\mathcal{M}^{(2)}_{i_1,\ldots,i_k}$) are needed to secure Gaussian cumulants[10]. These properties hold whenever the saddle point approximation of Eq.(66-67) is valid. In such cases one expects Gaussian confidence levels for the Quasi-Monte Carlo estimator E_1.

2 Multi-point distributions with diaphonies

2.1 Diaphony

We consider a point set X with N elements, given in C. The non-uniformity of the point set X can be described by its *diaphony*[11]:

$$D(X) = \frac{1}{N}\sum_{j,k=1}^{N} \beta(\vec{x}_j, \vec{x}_k) \quad , \tag{40}$$

with

$$\begin{aligned} \beta(\vec{x}_j, \vec{x}_k) &= \hat{\sum_{\vec{n}}} \sigma^2_{\vec{n}}\, e_{\vec{n}}(\vec{x}_j)\bar{e}_{\vec{n}}(\vec{x}_k) \quad , \\ e_{\vec{n}}(\vec{x}) &= \exp(2i\pi\, \vec{n}\cdot\vec{x}) \quad . \end{aligned} \tag{41}$$

Here, the vectors $\vec{n} = (n_1, n_2, \ldots, n_d)$ form the integer lattice, and the hat denotes the sum over all $\vec{n}$ except $\vec{n} = \vec{0}$. We may also write

$$D(X) = \frac{1}{N}\hat{\sum_{\vec{n}}} \sigma^2_{\vec{n}} \left|\sum_{j=1}^{N} e_{\vec{n}}(\vec{x}_j)\right|^2 \quad , \tag{42}$$

so that we recognize the diaphony as a measure of how well the various Fourier modes are integrated by the point set X. The diaphony is therefore seen to be related to the 'spectral test', well-known in the field of random-number generator testing. For the *mode strengths* $\sigma^2_{\vec{n}}$ we have

$$\sigma^2_{\vec{n}} \geq 0 \quad , \quad \hat{\sum_{\vec{n}}} \sigma^2_{\vec{n}} = 1 \quad . \tag{43}$$

[10] Approach to a Gaussian distribution,for iid random variables, would require $c_n/(c_2)^{n/2}$ to approach 0 for large N.

[11] some of the concepts of this section have also been discussed in [2] and [3].

The latter convention simply establishes the overall normalization of D. The advantage of this diaphony over, say, the usual (star)discrepancy is the fact that it is translation-invariant:

$$\beta(\vec{x}_j, \vec{x}_k) = \beta(\vec{x}_j - \vec{x}_k) \ , \tag{44}$$

so that point sets X and X' that differ only by a translation (modulo 1) have the same non-uniformity: the diaphony is actually defined on the hyper-torus rather than on the hypercube. Also, the diaphony is *tadpole-free*:

$$\int_C \beta(\vec{x}) \ d^d\vec{x} = 0 \ . \tag{45}$$

Moreover, we shall use $\sigma^2_{\vec{n}}$ such that $\sigma^2_{\vec{n}} = \sigma^2_{\vec{n}'}$ if the two lattice vectors $\vec{n}$ and $\vec{n}'$ differ only by a permutation of their components. Thus, X and X' will also have the same non-uniformity if they differ by a global permutation of the coordinates of the points.

2.1.1 Some numerical results

In this section the behavior of a specific diaphony is presented, for three point sequences, as the number of points N increases.

The diaphony is defined by Eq.(42) with

$$\sigma_{\vec{n}} = K e^{-\lambda \vec{n}^2} \quad K^{-1} = \sum e^{-\lambda \vec{n}^2} \quad \lambda = 0.1 \tag{46}$$

The reason for experimenting with this definition lies in the factorizing property of the $\sigma_{\vec{n}}$. Due to K^{-1} being related to Jacobi theta functions, we call this the 'Jacobi diaphony'. We will be using this diaphony in most of what follows.

In this paper we will be using three point sequences that we will be calling `Ranlux`, `Van Der Corput` and `Niederreiter`. `Ranlux` is a pseudo-random point sequence generated by the `Ranlux` algorithm (see [4]) with luxury level equal to 3. `Van Der Corput` is a quasi-random sequence generated by an implementation of the algorithm by Halton that generalizes to many dimensions an older algorithm by Van der Corput (see [5]) with prime bases $2, 3, 5, 7, 11, \ldots$. Finally `Niederreiter` is another, optimal[12] quasi-random sequence based on the algorithm in (see [7]). In particular, we follow the choices of [8] and construct the sequence in whichever base is optimal for the current dimension (see [7]).

[12] in a sense described in [7] and [8].

It should also be noted that only modes with square length up to $\vec{n}^2 \leq 15$ are included in the calculation of the diaphony (including the determination of K), in anticipation of the same restriction on the estimator sums in later sections.

In the plots that follow, the diaphony of the `Niederreiter` sequence in particular, but also that of the `van der Corput` sequence, exhibited a large variation in relatively small intervals of N. As the number of points N approaches certain critical values the diaphony reaches very small levels, only to return to its 'cruising' values a few points later. To avoid cluttering the plots we present here the diaphony averaged in packs of 500 points without information on the minimum or maximum value found in each pack. The minimum values for each pack, that correspond to exceptional point configurations, are very interesting on their own but do not affect the present study.

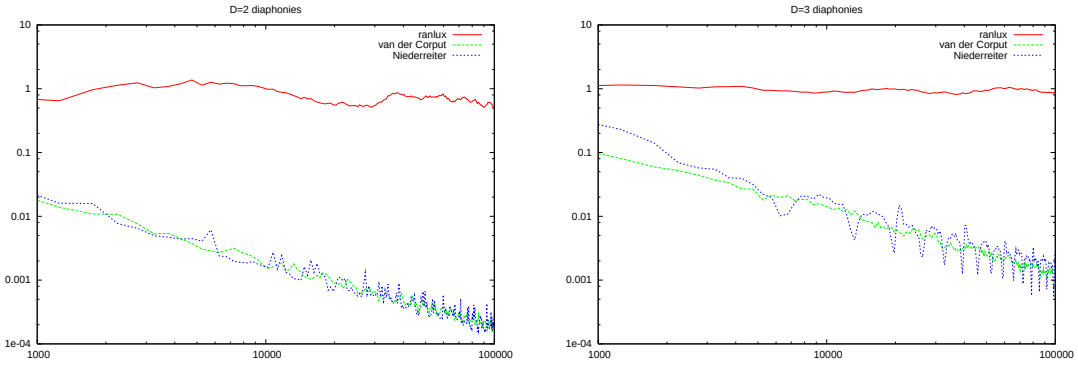


Figure 1: D=2 (left) and D=3 (right). The diaphony of `RANLUX` (red line), `Van Der Corput` (green line) and `Niederreiter` (blue line).

The diaphony of the `RANLUX` sequence is seen to oscillate around 1, as expected. Moreover the behavior of the `Niederreiter` sequence improves with the number of dimensions when compared with crude `Van der Corput`, an encouraging hint for higher dimensions.

2.2 Generating function

We shall now compute a $1/N$ approximation to the moment-generating distribution of the p-point probability distribution, that is,

$$G_p(z) = \left\langle \exp(zD(X)) \right\rangle_{\vec{x}_{p+1},\vec{x}_{p+2},\ldots,\vec{x}_N} \quad , \tag{47}$$

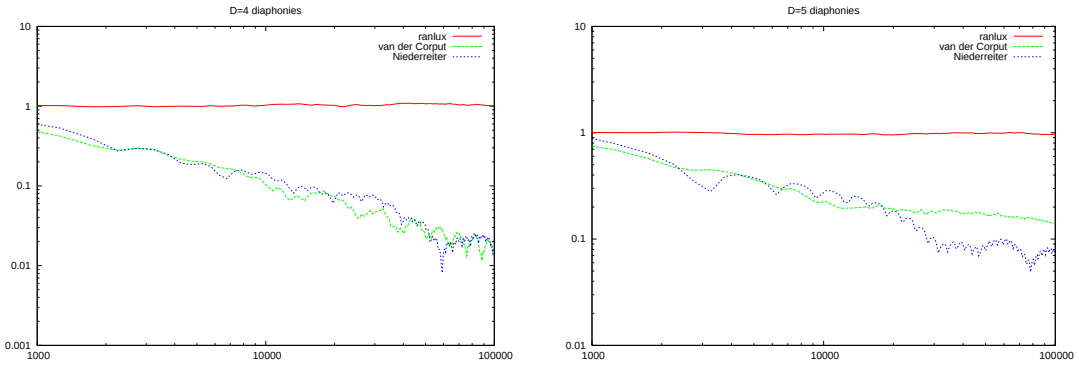


Figure 2: D=4 (left) and D=5 (right). The diaphony of `RANLUX` (red line), `Van Der Corput` (green line) and `Niederreiter` (blue line).

where we have indicated that the points $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_p$ are kept fixed while the remaining $N - p$ points are integrated over. $G_p(z)$ therefore still depends on $\vec{x}_1, \ldots, \vec{x}_p$. This is most easily achieved using a diagrammatic approach, which has been introduced in [2]. We shall indicate with crosses those points that are kept fixed (with an implied sum over them, from 1 to p), and with dots ('beads') those points that are integrated over (again, with an implied sum running from $p+1$ to N). The function β is indicated by a solid line. As the simplest examples, then, we have

$$\text{if } p = N\text{:} \quad \frac{1}{N} \times\!\!\longrightarrow\!\!\times = \frac{1}{N} \sum_{j,k=1}^{N} \beta(\vec{x}_j - \vec{x}_k) = D(X) \quad , \tag{48}$$

and

$$\text{if } p = 0\text{:} \quad \langle D(X) \rangle_{\vec{x}_1,\ldots,\vec{x}_N} = \beta(0) = \bigcirc = 1 \quad . \tag{49}$$

Other examples are

$$\bigcirc \;\; = \int_C \beta(\vec{x}_1 - \vec{x}_2)^2 \, d^d\vec{x}_1 \, d^d\vec{x}_2 \quad ,$$

$$\bigcirc \;\; = \int_C \beta(\vec{x}_1 - \vec{x}_2)\beta(\vec{x}_2 - \vec{x}_3)\beta(\vec{x}_3 - \vec{x}_1) \, d^d\vec{x}_1 \, d^d\vec{x}_2 \, d^d\vec{x}_3 \quad , \tag{50}$$

and so on: a general closed loop with precisely n beads will be denoted by $\textcircled{n}$. Note that, since, the functions $e_{\vec{n}}(\vec{x})$ form an orthonormal (and even a complete)

set, we have

$$\textcircled{n} = \sum_{\vec{n}}^{\wedge} \left(\sigma_{\vec{n}}^2\right)^n \quad . \tag{51}$$

We can now simply write out all possible (connected and disconnected) diagrams where every solid line ends in a cross or a bead, and apply the following Feynman rules:

1. A factor $2z/N$ for every β line (where the factor 2 arises from the two possible orientations);
2. A factor $(N-p)^{\underline{q}}$ for every diagram (or product of diagrams) that contains precisely q beads[13];
3. In addition, the usual symmetry factors arising from equivalent lines and vertices, and from the repetition of identical (sub)diagrams.

We shall compute $G_p(z)$ including terms of order 1 and those of order $1/N$. Note that

$$(N-p)^{\underline{q}} = N^q \left(1 - \frac{pq}{N} - \frac{q(q-1)}{2N}\right) + \mathcal{O}(N^{-2}) \tag{52}$$

as long as $N \gg pq, q^2$. In the following we shall always assume this.

First, we consider contributions without any crosses or nontrivial vertices. A general term in this class is given by

$$\frac{(N-p)^{\underline{Q}}}{N^Q} \frac{1}{r_1!} \left(z \bigcirc\right)^{r_1} \frac{1}{r_2!} \left(z^2 \bigcirc\right)^{r_2} \frac{1}{r_3!} \left(\frac{4z^3}{3} \bigcirc\right)^{r_3} \cdots \quad ,$$

where

$$Q = r_1 + 2r_2 + 3r_3 + \cdots \quad ; \tag{53}$$

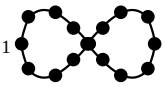
up to order $1/N^2$, this contribution to the generating function can therefore be written as

$$\begin{aligned} G_p^{(1)}(z) &= \left(1 - \frac{pz}{N}\frac{\partial}{\partial z} - \frac{z^2}{2N}\frac{\partial^2}{\partial z^2}\right) \sum_{\{r\}} \prod_{n\geq 1} \frac{1}{r_n!} \left(\frac{(2z)^n}{2n} \textcircled{n}\right)^{r_n} \\ &= \left(1 - \frac{pz}{N}\frac{\partial}{\partial z} - \frac{z^2}{2N}\frac{\partial^2}{\partial z^2}\right) G^{(0)}(z) \quad , \\ G^{(0)}(z) &= \exp\left(-\frac{1}{2}\sum_{\vec{n}}^{\wedge} \log\left(1 - 2z\sigma_{\vec{n}}^2\right)\right) \quad . \end{aligned} \tag{54}$$

[13] The 'falling power' is defined as $a^{\underline{b}} = a!/(a-b)! = a(a-1)(a-2)\cdots(a-b+1)$.

Up to $1/N^2$, one diagram with a four-point vertex may be present: a generic contribution of this type is

$$\frac{(N-p)^{\underline{Q+m+1+m_2+1}}}{N^{Q+m_1+m_2+2}}\left(\frac{(2z)^{m_1+m_2+2}}{8}\right.$$



$$\left.\times\frac{1}{r_1!}\left(z\right.\right.$$

$$\left.\right)^{r_1}\frac{1}{r_2!}\left(z^2\right.$$

$$\left.\right)^{r_2}\frac{1}{r_3!}\left(\frac{4z^3}{3}\right.$$

$$\left.\right)^{r_3}\cdots\ ,$$

where $m_{1,2}$ denote the number of beads on each loop, excluding the one on the four-vertex. Let us define

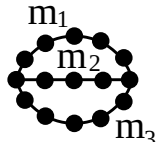
$$\phi(z;\vec{x}_j-\vec{x}_k)=\sum_{\vec{n}}^{\widehat{}}\frac{2z\sigma_{\vec{n}}^2}{1-2z\sigma_{\vec{n}}^2}e_{\vec{n}}(\vec{x}_j)\bar{e}_{\vec{n}}(\vec{x}_k)\ \ ; \tag{55}$$

then, this contribution can be written as

$$G_p^{(2)}(z)=\frac{1}{8N}\phi(z;0)^2G^{(0)}(z)\ \ . \tag{56}$$

Note that the lemniscate graph is actually equal to the product of two closed loops: this is a consequence of the translational invariance of the diaphony. A generic contribution containing two three-vertices is

$$\frac{(N-p)^{\underline{Q+m_1+m_2+m_3+2}}}{N^{Q+m_1+m_2+m_3+3}}\left(\frac{(2z)^{m_1+m_2+m_3+3}}{12}\right.$$



$$\left.\times\frac{1}{r_1!}\left(z\right.\right.$$

$$\left.\right)^{r_1}\frac{1}{r_2!}\left(z^2\right.$$

$$\left.\right)^{r_2}\frac{1}{r_3!}\left(\frac{4z^3}{3}\right.$$

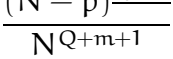
$$\left.\right)^{r_3}\cdots\ ,$$

so that this contribution to the generating function reads

$$G_p^{(3)}(z)=\frac{1}{12N}G^{(0)}(z)\int_C\phi(z;\vec{x})^3\,d^d\vec{x}\ \ . \tag{57}$$

The diagrams with crosses have the generic contribution

$$\frac{(N-p)^{\underline{Q+m}}}{N^{Q+m+1}}\left(z(2z)^m\right.$$



$$\left.\right)$$

$$\times \frac{1}{r_1!}\left(z\ \text{  }\right)^{r_1} \frac{1}{r_2!}\left(z^2\ \text{}\right)^{r_2} \frac{1}{r_3!}\left(\frac{4z^3}{3}\ \text{}\right)^{r_3} \cdots \ ,$$

leading to

$$\begin{aligned} G_p^{(4)}(z) &= \frac{1}{2N} G^{(0)}(z) \sum_{j,k=1}^{p} \phi(z; \vec{x}_j - \vec{x}_k) \\ &= \frac{1}{N} G^{(0)}(z) \left(\frac{p}{2} \phi(z;0)^2 + \sum_{1 \le j < k \le p} \phi(z; \vec{x}_j - \vec{x}_k) \right) \quad , \end{aligned} \tag{58}$$

where we have singled out the contributions with $j = k$. All other possible diagrams either vanish because of translational invariance and tadpole-freedom, or are of order $1/N^2$ or lower. The final result for the generating function up to order $1/N^2$ is therefore

$$\begin{aligned} G_p(z) &= G^{(0)}(z) \left(1 - \frac{1}{4N} \int_C \phi(z; \vec{x})^2 \, d^d\vec{x} + \frac{1}{12N} \int_C \phi(z; \vec{x})^3 \, d^d\vec{x} \right. \\ &\qquad \left. + \frac{1}{N} \sum_{1 \le j < k \le p} \phi(z; \vec{x}_j - \vec{x}_k) \right) \quad . \end{aligned} \tag{59}$$

Note that the term in $G^{(1)}(z)$ containing p cancels precisely against that in $G^{(4)}(z)$, so that the only reference to p is in the last term in brackets in Eq.(59), and indeed we have

$$\int_C G_p(z) \, d^d\vec{x}_p = G_{p-1}(z) \quad . \tag{60}$$

In Appendix B we give the result for the higher order ($O(\frac{1}{N^2})$) term in G_p. There are 25 terms that contribute but only three of them include p. The condition 60 still holds.

2.3 Multi-point distribution by Laplace transform

From the generating function, we can recover the actual probability distributions. As discussed above, let $H(s)$ be the probability that the point set X has diaphony equal to s, that is, $D(X) = s$. The underlying ensemble of point sets is that of sets

of N iid uniformly distributed points, *i.e.* the same ensemble underlying the usual Monte Carlo error estimates. Then, we have

$$\begin{aligned} H(s) &= \int_C d^d\vec{x}_1\ d^d\vec{x}_2 \cdots d^d\vec{x}_N\ \delta(D(X) - s) \\ &= \frac{1}{2i\pi} \int_{-i\infty}^{+i\infty} e^{-zs}\ G_0(z)\ dz\ , \end{aligned} \tag{61}$$

where the integration contour runs to the left of all the singularities of $G_0(z)$; and the multi-point distribution for p points averaged over all point sets X with diaphony s, is given by

$$\begin{aligned} P_p(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) &= \frac{1}{H(s)} R_p(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p)\ , \\ R_p(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) &= \frac{1}{2i\pi} \int_{-i\infty}^{+i\infty} e^{-zs}\ G_p(z)\ dz\ . \end{aligned} \tag{62}$$

Since we write the deviation from uniformity of the multi-point distribution as

$$P_p(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) = 1 - \frac{1}{N} F_p(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p)\ , \tag{63}$$

we see that the multi-point correlation F_p is, up to $O(\frac{1}{N})$, as claimed, built up from two-point correlators[14]: for $p \geq 3$,

$$F_p(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_p) = F_{p-1}(s;\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_{p-1}) + \sum_{j=1}^{p-1} F_2(s;\vec{x}_j,\vec{x}_p)\ , \tag{64}$$

so that the p-point correlator is simply the sum of all $p(p-1)/2$ 2-point correlators. In the approximation used, the sub-leading terms in $H(s)$ are actually irrelevant, and we may write

$$H(s) \approx \frac{1}{2i\pi} \int_{-i\infty}^{+i\infty} \exp(\psi(s;z))\ dz\ ,$$

[14]This doesn't hold for the next order in $\frac{1}{N}$ as seen in Appendix B. Terms like the one of Eq.(109), that don't factorize, appear for $p \geq 3$.

$$\begin{aligned} \psi(s;z) &= -sz - \frac{1}{2}\hat{\sum_{\vec{n}}} \log\left(1 - 2z\sigma_{\vec{n}}^2\right) \quad , \\ F_2(s;\vec{x}_1,\vec{x}_2) &= \frac{-1}{2\pi i H(s)} \int_{-i\infty}^{+i\infty} \exp(\psi(s;z))\phi(z;\vec{x}_1-\vec{x}_2)\ dz \quad . \end{aligned} \tag{65}$$

Except in the very simplest cases[15], a complete evaluation of Eq.(65) is nontrivial. A simplification arises if s is much smaller than its expectation value 1 (which is anyway the aim in quasi-Monte Carlo), or if the Gaussian limit is applicable, namely when the number of modes with non-negligible $\sigma_{\vec{n}}^2$ becomes large in such a way that no single mode dominates. In practice, this happens when the dimensionality of C becomes large. Fortunately, these are precisely the situations of interest. The position of the saddle point for $H(s)$, $\hat{z}$, is given by

$$\hat{\sum_{\vec{n}}} \frac{\sigma_{\vec{n}}^2}{1 - 2\hat{z}\sigma_{\vec{n}}^2} = s \quad . \tag{66}$$

For $s \ll 1$, therefore, $\hat{z}$ is large and negative. Since to first order the same saddle point may be used for R_2, we find the attractive result

$$F_2(s;\vec{x}_1,\vec{x}_2) \approx \hat{\sum_{\vec{n}}} \omega_{\vec{n}}\ e_{\vec{n}}(\vec{x}_1)\ \bar{e}_{\vec{n}}(\vec{x}_2) \quad , \quad \omega_{\vec{n}} = \frac{2\hat{z}\sigma_{\vec{n}}^2}{2\hat{z}\sigma_{\vec{n}}^2 - 1} \quad . \tag{67}$$

The formulae (66) and (67) suffice, in our approximation, to compute all the multipoint correlations.

We finish this section with the following observation. Suppose that F_2 is given as a function of $\vec{x}_1, \vec{x}_2$. By Fourier integration we can then compute the $\omega_{\vec{n}}$. The assumption that the saddle-point approximation is valid, together with the normalization condition $\sum \sigma_{\vec{n}}^2 = 1$, then allows us to write

$$\hat{z} = -\frac{1}{2}\hat{\sum_{\vec{n}}} \frac{\omega_{\vec{n}}}{1-\omega_{\vec{n}}} \quad , \quad \sigma_{\vec{n}}^2 = -\frac{1}{2\hat{z}}\frac{\omega_{\vec{n}}}{1-\omega_{\vec{n}}} \quad , \quad s = \hat{\sum_{\vec{n}}} \sigma_{\vec{n}}^2(1-\omega_{\vec{n}}) \quad . \tag{68}$$

We see that F_2 not only determines the *form* of the diaphony, but in addition also its *value*.

[15] See section 4.3.

3 Application of Quasi-Monte Carlo estimators

3.1 The mechanism behind error reduction

After the above preliminaries we can now examine the mechanism by which Quasi-Monte Carlo can outdo Monte Carlo. We shall assume the saddle-point approximation to be valid. For $s < 1$, we then have $\hat{z} < 0$, and all the $\omega_{\vec{n}}$ are positive, and as $\hat{z} \to -\infty$ they approach unity from below (although for $|\vec{n}| \to \infty$ they must always, of course, go to zero). Now notice that the set of functions $e_{\vec{n}}(\vec{x})$ is complete, that is,

$$\sum_{\vec{n}} e_{\vec{n}}(\vec{x}_1)\ \bar{e}_{\vec{n}}(\vec{x}_2) = \delta^d(\vec{x}_1 - \vec{x}_2)\ . \tag{69}$$

This allows us to write the variance of the Monte Carlo error as

$$\sigma\,(E_1)^2 = \frac{1}{N} \sum_{\vec{n}}^{\hat{}} \left| \int_C f(\vec{x})\ e_{\vec{n}}(\vec{x})\ d^d\vec{x} \right|^2 , \tag{70}$$

where the contribution from the zero mode $\vec{n} = 0$ is canceled by the J_1^2 term. For Quasi-Monte Carlo on the other hand, we find

$$\sigma\left(E_1^{(q)}\right)^2_{(q)} = \frac{1}{N} \sum_{\vec{n}}^{\hat{}} (1 - \omega_{\vec{n}}) \left| \int_C f(\vec{x})\ e_{\vec{n}}(\vec{x})\ d^d\vec{x} \right|^2 . \tag{71}$$

We see that those modes $\vec{n}$ for which $\omega_{\vec{n}}$ is positive tend to lead to an error reduction. In the saddle-point approximation, therefore, *any* value $0 < s < 1$ will lead to a decreased error with respect to standard Monte Carlo. On the other hand, since

$$0 < \hat{z} < \min_{\vec{n}} \frac{1}{2\sigma^2_{\vec{n}}} \quad \text{for } s > 1\ , \tag{72}$$

large values of the diaphony will actually lead to an *increase* in the error. Note that in the above we have only used the fact that the $e_{\vec{n}}$ form a *complete, orthonormal* set of functions: therefore, the error-reduction result holds for a much wider class of discrepancies than just the diaphonies discussed in this paper.

3.2 Estimators analyzed

We can now arrive at an estimator for the Quasi-Monte Carlo error. The simplest form is obtained by inserting Eq.(67) in the equation for E_2 (Eq.26):

$$E_2^{(q)} = \frac{1}{N^2}\sum f_i^2 - \frac{1}{N^3}(\sum f_i)^2 - \frac{1}{N^3}\sum_{\hat{\vec{n}}} \omega_{\vec{n}} |\sum_i f_i e_{\vec{n}}(x_i)|^2 \qquad (73)$$

with

$$\omega_{\vec{n}} = \frac{-2\hat{z}\sigma_{\vec{n}}^2}{1 - 2\hat{z}\sigma_{\vec{n}}^2} \qquad (74)$$

We are still free to choose the exact form of the weights $\sigma_{\vec{n}}^2$ at will, under the constraints of Eq.(43). Our choice is the so called Jacobi weights[16]

$$\sigma_{\vec{n}}^2 = K e^{-\lambda \vec{n}^2} \qquad (75)$$

with

$$K^{-1} = \hat{\sum_{\vec{n}}} e^{-\lambda \vec{n}^2} \qquad (76)$$

The parameter λ controls the 'sensitivity' of the diaphony: as $\lambda \to 0$ we get $\sigma_{\vec{n}} \to 1$ for every mode which corresponds to a super-sensitive diaphony, useless for practical purposes, while as $\lambda \to \infty$ only the modes with $\vec{n}^2 = 1$ contribute making the diaphony fairly non-sensitive. We choose $\lambda = 0.1$. Other values of λ, within a 'reasonable range' do not alter, in practice, the numerical value of $E_2^{(q)}$, as shown in section 4.1.

It is easy to see that the estimator averages (to leading order in N) in a positive definite quantity [17]. This leaves still open the possibility for a negative error estimate, particularly for relatively smooth functions where the cancellation between the two sums of the pseudo estimate are large leading to a small error. The source of the negative error effect is clear in the case of a constant function. Then

$$f(x) = C \Rightarrow E_2^{(q)} = -\frac{1}{N^3} C^2 \sum_{\vec{n}} \omega_{\vec{n}} \sum_{i,j} u_{\vec{n}}(x_i)\bar{u}_{\vec{n}}(x_j) \qquad (77)$$

and the point sum of every Fourier mode can be anything from 0 (when the points are spread evenly enough to produce complete cancellations for all the included modes) to N^2 (when all the points are on top of each other). The average of this

[16] due to their convenient factorizing property.

[17] It averages to Eq.(71) which is positive definite as long as $s < 1$.

sum is N (for truly random points), but for Quasi-Monte Carlo points we expect that this sum will be significantly smaller than that. For non-constant functions similar effects can be expected, apart from the fact that the first two terms of $E_2^{(q)}$ do not cancel anymore. Thus, we expect negative squared errors for higher modes or small number of points, and this is what has been observed in a number of plots. Unfortunately there is no way to predict precisely when, as N increases, the estimator gets a useful, positive value. One could resort to the error of $E_2^{(q)}$, but that is cubic in the number of modes (see Eq.29) and hence prohibitively expensive in realistic calculations.

The way out of this is the estimator of Eq.(31) which can be written in a form with unrestricted sums as follows:

$$\begin{aligned} E_2^{(q)} &= \frac{1}{N^{\underline{2}}}S_2 - \frac{1}{NN^{\underline{2}}}S_1^2 - \frac{(N-1)^{\underline{3}}}{NN^{\underline{4}}}\sum_{\vec{n}}\omega_{\vec{n}}|W_{\vec{n}}|^2 \\ &+\frac{(N-1)^{\underline{3}}}{NN^{\underline{4}}}S_2\sum_{\vec{n}}\omega_{\vec{n}} + \frac{N-1}{NN^{\underline{4}}}\sum_{\vec{n}}\omega_n(2S_1\Re\left\{W_{\vec{n}}\bar{U}_{\vec{n}}\right\} - 2\Re\left\{U_{\vec{n}}\bar{Q}_{\vec{n}}\right\}) \\ &-\frac{1}{NN^{\underline{4}}}\sum_{\vec{n}}\omega_{\vec{n}}(N-2+|U_{\vec{n}}|^2)(S_2 - S_1^2) \end{aligned} \tag{78}$$

where

$$\begin{aligned} U_{\vec{n}} &\equiv \sum_i u_{\vec{n}}(x_i) \\ W_{\vec{n}} &\equiv \sum_i u_{\vec{n}}(x_i)f(x_i) \\ Q_{\vec{n}} &\equiv \sum_i u_{\vec{n}}(x_i)f^2(x_i) \\ S_2 &\equiv \sum_i f_i^2 \\ S_1 &\equiv \sum_i f_i \end{aligned} \tag{79}$$

It is identically zero for a constant function, as can be easily checked, and averages to the leading order of the squared variance of E_1. The correction terms are of higher than leading order in N, but that does *not* mean that we have selectively included some NLO corrections to the variance. The correction terms above are such that the NLO terms vanish on the average.

In practice the infinite sum over modes in both estimators has to be truncated. This should not be perceived as an approximation of any kind. It amounts to a redefinition of the diaphony. Looking at Eq.(66) we see that as the value of s becomes small the saddle point becomes quickly large and negative: $\hat{z} \ll 0$. Then $-2\hat{z}\sigma^2_{\vec{n}} \to \infty$ for low modes and $-2\hat{z}\sigma^2_{\vec{n}} \to 0$ for higher modes, when $\sigma^2_{\vec{n}}/|\hat{z}| \to 0$. We can, thus, safely neglect these higher modes in the estimator. As long as the value of the diaphony is small, which is in any case the goal in Quasi--Monte Carlo , the profile of $\omega_{\vec{n}}$ depends only on the choice of λ, which, as said, also regulates the sensitivity of the diaphony. We see therefore that the estimator inherits the sensitivity of the diaphony in a direct way.

It is worth noting that the factorized form of the β-function in the diaphony definition is directly responsible for the fact that the two estimators are now of complexity $N \times M$ (with M the number of modes) instead of quadratic in N. This is a desirable achievement as long as $M \leq N$, which we shall always assume to be the case.

3.3 Numerical results

In the following we will present a number of plots that show how both the 'classical' and the quasi error estimates[18] behave as a function of the number of points N. In the process we will use the three types of point sequences defined in section 2.1.1.

A number of test functions were used for integrands. They consist of a subset of the test functions used by Schlier in [6], along with a Gaussian function with dimension-dependent width. We have

$$TF13 : f(\vec{x}) = \prod_{k=1}^{D} \frac{|4x_k - 2| + k}{1 + k} \tag{80}$$

which averages to $J_1 = 1$. This test function is especially tailored for a Van der Corput sequence, since in $D = 1$ it is perfectly integrated by such a sequence with base 2.

$$TF2 : f(\vec{x}) = \prod_{k=1}^{D} k\cos(kx_k) \tag{81}$$

[18]The 'classical' or 'pseudo' estimator, E_2, is the one of Eq.(7), constructed on the assumption that the points are iid. By 'quasi' estimator, $E_2^{(q)}$, we mean the 'improved' estimator Eq.(78).

which averages to $J_1 = \prod_k \sin(k)$. This function should be difficult to integrate in high dimensions.

$$TF4 : f(\vec{x}) = \sum_{k=1}^{D} \prod_{j=1}^{k} x_j \tag{82}$$

which averages to $J_1 = 1 - \frac{1}{2^D}$. It is chosen as a simple example of a function that is not a product of single-variable functions.

A Gaussian with fixed width suffers from a rapid decrease, in higher dimensions, of the region of the integration volume where the function is non-zero, making the integration cumbersome (the higher the dimension, the more points are needed and inter-dimensional comparison is difficult). To avoid this we use instead

$$TF6 : f(\vec{x}) = \prod_{i=1}^{D} \sum_{n_i=-\infty}^{\infty} \frac{e^{-(x_i - x_{0i} + n_i)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \tag{83}$$

which is a product of superpositions of a Gaussian and its tails outside the $[0, 1]$ interval. We wish to keep the variance of this function independent of the number of dimensions, so we define σ such that

$$\frac{1}{2\sigma} \sum_{m=-\infty}^{\infty} e^{-m^2/4\sigma^2} = (1 + V)^{1/D} \sqrt{\pi} \tag{84}$$

where in practice it suffices to keep the first couple of terms in the sum. The function averages to $J_1 = 1$ and spreads as the number of dimension grows ($\sigma \to \infty$ as $D \to \infty$).

In the following plots the error and its estimates as functions of the number of points N are shown in a double logarithmic scale.

The 'classical' error estimate is presented, along with three versions of quasi error estimators, $E_2^{q5}, E_2^{q10}, E_2^{q15}$. The superscript next to q denotes the squared length of the highest modes included in the sums of Eq.(78). Thus E_2^{q10} includes[19] modes with $\vec{n}^2 \leq 10$. In table 1 we give the number of modes with $\vec{n}^2 \leq 15$, and $\vec{n} \leq 5$ for different dimensions. It is evident that the number of modes grows rapidly with the dimensionality.

[19] please note that the square length of a mode is the sum of the squares of D integers. So for $D = 2$, for example, the modes present are those with square equal to $1, 2, 4, 5, 9, 10, 13, 16, 17, \ldots$ and, thus, $E_2^{(q15)}$ actually contains modes with squared length up to 13.

D	# of modes
1	6
2	44
3	250
4	1256
5	5182

D	# of modes
1	4
2	20
3	56
4	136
5	332

Table 1: number of modes with $\vec{n}^2 \leq 15$ (left) and $\vec{n}^2 \leq 5$ (right)

The real error made is included for comparison. The data were collected in a point per point basis up to $N = 10^5$. In the plots we have included the average value of each error for successive subsets of 500 points, suppressing any information on minimum or maximum values in the subset[20].

All integrations are performed in the unit hypercube $[0,1]^D$. The dimensionality varies from 2 to 6.

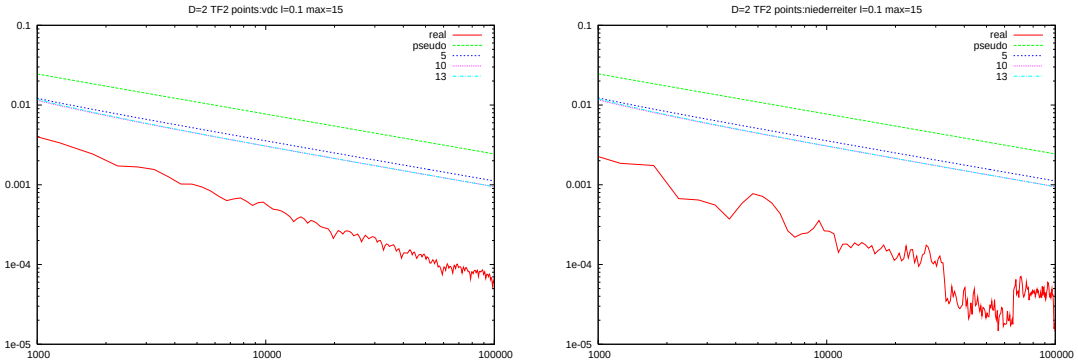


Figure 3: TF2, d=2 log-plot using a `Van der Corput` sequence (left) and a `Niederreiter` sequence (right). The classical error estimator is far off the real error whereas the quasi estimators are approaching the real error as more modes are added to the sum. The need for more modes is, however, obvious, in both plots.

[20]The real error (in particular) fluctuates a lot as the quasi sets complete their successive cycles of low diaphony, but knowledge of the specific point where the error minimizes is of course not available a priori.

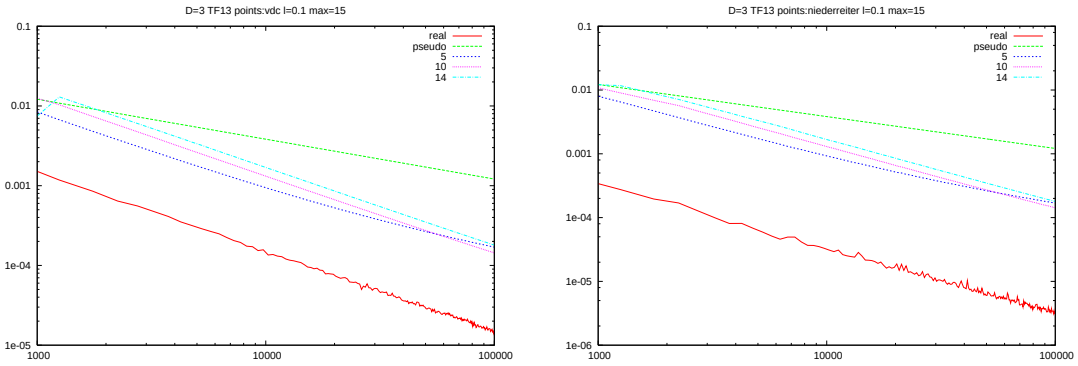


Figure 4: TF13, d=3 log-plot using a `Van der Corput` sequence (left) and a `Niederreiter` sequence (right). The quasi estimators follow the error with the appropriate N-dependence contrary to the pseudo estimator. Note that the E^{q14} is in this case worse than E^{q10} or E^{q5} for all $N \leq 100000$. The higher modes converge slower to their average value, but the cross-over point is not known in advance and it is function-dependent.

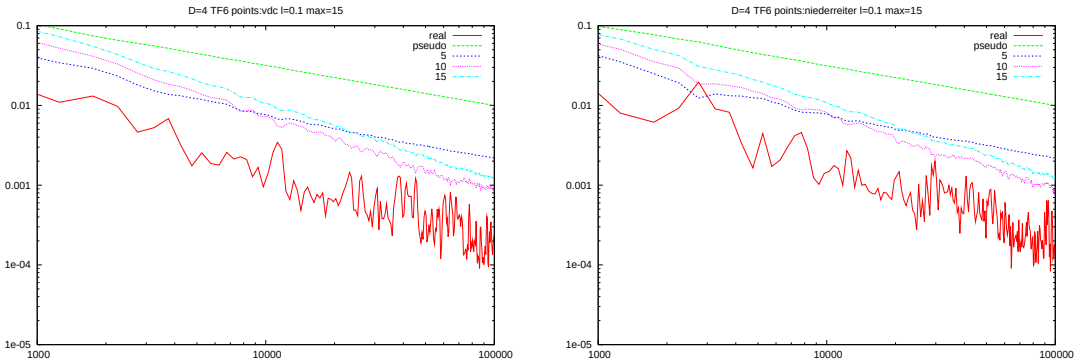


Figure 5: TF6, d=4 log-plot using a `Van der Corput` sequence (left) and a `Niederreiter` sequence (right). The quasi estimators approximate well the error. Moreover we see here a clearer instance of the crossover of higher modes in large N mentioned in the previous figure.

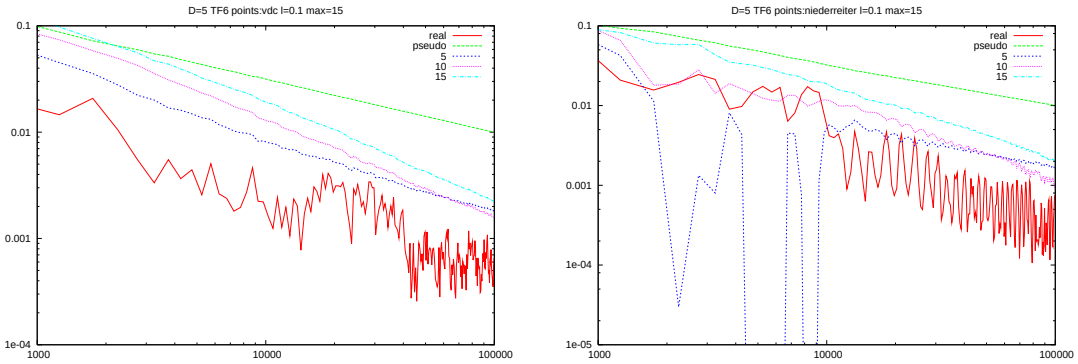


Figure 6: TF6, d=5 log-plot using a `Van der Corput` sequence (left) and a `Niederreiter` sequence (right). The use of the improved estimator (Eq.78) reduces the probability of a negative error square estimate but, naturally, it doesn't remove it altogether. The plot on the right demonstrates this effect. As expected, the estimator returns to positive values and stabilizes as the number of points increases and the estimator converges to its average value.

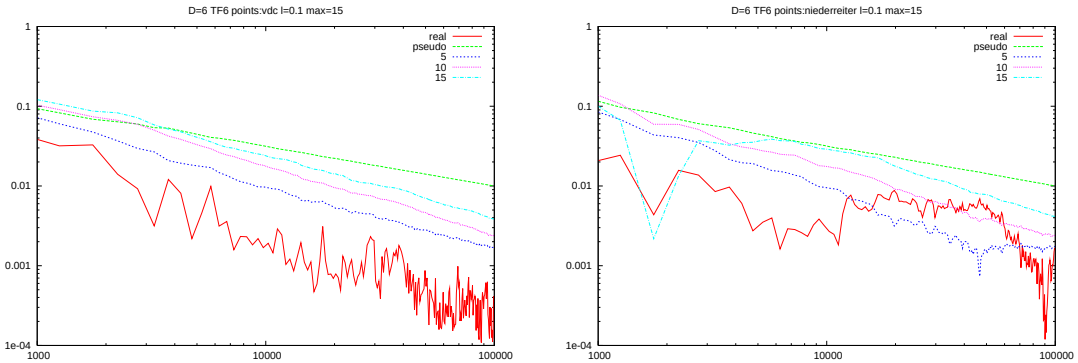


Figure 7: TF6, d=6 log-plot using a `Van der Corput` sequence (left) and a `Niederreiter` sequence (right). In this case the estimators describe very well the real error made in the integration.

4 Alternative approaches

4.1 Raising the value of λ in the Jacobi diaphony

In general the real Quasi-Monte Carlo error is approached by including more and more modes in the estimator sum. At the same time, by including higher modes, one increases the error on this estimate (the error on E_2) because one attempts to estimate by Monte Carlo means the integral $\int f(\vec{x}) e_{\vec{n}}(\vec{x})$ which will fluctuate vigorously for higher modes.

One might then attempt to raise the value of λ, thus decreasing the number of active modes (that give an appreciably non-zero $\omega_{\vec{n}}$)). This would of course reduce the sensitivity of the diaphony, artificially lowering its value. Improvement in the error estimate originating from higher modes would be lost but the contribution of the modes close to the origin (which are the ones included) would be relatively enhanced, as can be seen from the behavior of the weights $\omega_{\vec{n}}$ (see Eq.67) .

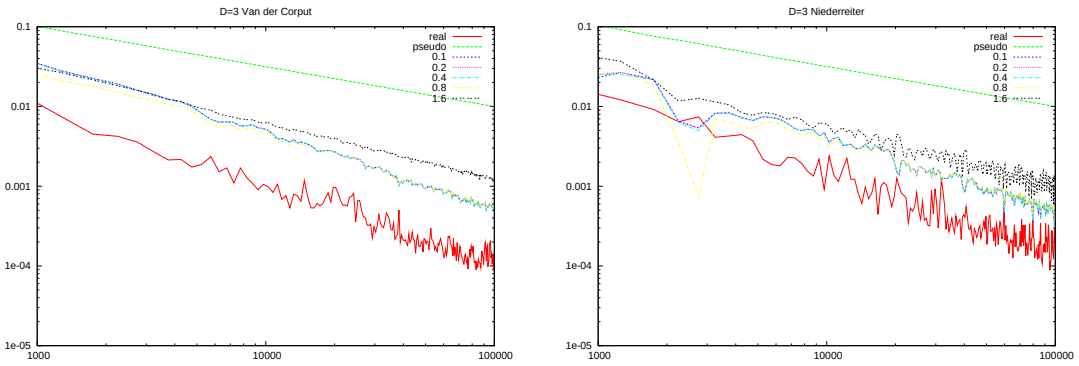


Figure 8: TF6, d=4 log-plot using the `Van der Corput` (left) and the `Niederreiter` (right) sequences. E^{q15} is shown for different values of λ indicated in the key, along with the real error and the classical estimate. Average values of all quantities for sets of 500 points are shown in each case. The value of λ doesn't alter the estimator, as long as that value stays within a specific range. We see that, in this case, the value $\lambda = 1.6$ is out of the safe range.

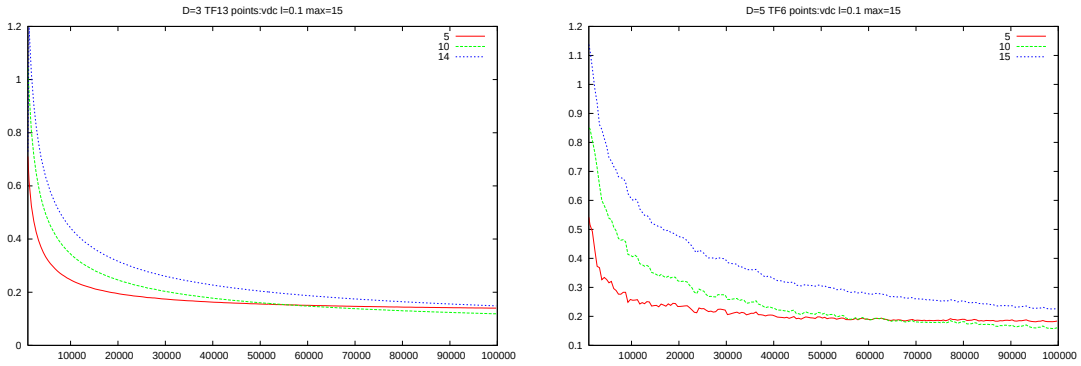


Figure 9: The ratio of different quasi estimators with the classical estimator for $d = 3$ and TF13 (left) and $d = 5$ TF6 (right). In both cases `Van der Corput` point-sets were used.

4.2 Monitored estimator

The estimators E_2^{q5} , E_2^{q10} and E_2^{q15} are not always proportional to the classical estimate, and, in some cases they decreases quite faster with N than the classical estimate does. They never decrease slower than the classical estimate, though, and one can use that as follows. One monitors the ratio of E_2^{q15}, for example, to the 'classical' error estimate, and after a certain point[21], the 'classical' error is only estimated and multiplied with that ratio. This is a purely linear algorithm and therefore very fast. Caution has to be exercised, though, in the way the critical ration is chosen, in order to avoid configurations where the estimators acquire a very low value for some exceptional value of N.

This approach relies heavily on the, frequently false, assumption that the quasi and classical estimators scale. If this is not so, the new estimate is conservative. One has, thus, the option to trade accuracy for cpu time.

The plots of fig.9 show the ratio of E_2^{q5} , E_2^{q10} and E_2^{q15} with E_2 for two particular cases.

[21] which depends on the resources of the user.

4.3 The box approximation

There is a choice for the diaphony that allows us to perform the integrals of Eq.(65) without resorting to the saddle point approximation. That choice is

$$\sigma_{\vec{n}}^2 = \frac{1}{M} \prod_{\mu=1..d} \theta(n^\mu \leq m) \tag{85}$$

for some arbitrary m. The normalization (Eq.43) determines $M = (2m+1)^D - 1$.

This diaphony includes only a finite number of modes, all of which are equally weighted. It can be seen as an approximation to the Jacobi diaphony since for small λ the latter gives $\sigma_{\vec{n}} \approx 1$ for $|\vec{n}| \leq n_c$ and $\sigma_{\vec{n}} \approx 0$ for $|\vec{n}| \geq n_c$ where n_c is determined implicitly by the value of the Jacobi diaphony. The diaphony can be evaluated as a quadratic function on the point-set from

$$S = \frac{1}{N} \sum_{\vec{n}} \sigma_{\vec{n}}^2 |\sum_i e_{\vec{n}}(\vec{x})|^2 = \frac{1}{NM} \sum_{|\vec{n}| \leq m} |\sum_i e_{\vec{n}}(\vec{x})|^2 \equiv \frac{1}{NM} \sum_{i,j} \psi(\vec{x}_i - \vec{x}_j) \tag{86}$$

with

$$\psi(\vec{x}_i - \vec{x}_j) = -1 + \prod_{\mu=1}^{D} \frac{\sin\left((2m+1)\pi(x_i^\mu - x_j^\mu)\right)}{\sin\left(\pi(x_i^\mu - x_j^\mu)\right)} \tag{87}$$

The distribution of point-sets with a particular value for s is then found by explicitly performing the z-integrals of Eq.(65):

$$H(s) = \frac{K^K s^{K-1}}{\Gamma(K)} e^{-Ks} \tag{88}$$

with $K \equiv M/2$. Hence the correlation function is

$$F(s; \vec{x}_i - \vec{x}_j) = \frac{(1-s)}{M} \psi(\vec{x}_i - \vec{x}_j) \tag{89}$$

and the estimator[22] of Eq.(73) becomes

$$E_2^{(q)} = \frac{1}{N^2} \sum f_i^2 - \frac{1}{N^3} (\sum f_i)^2 - \frac{1}{N^3} \frac{(1-s)}{M} \sum_{i,j} \psi(\vec{x}_i - \vec{x}_j) f_i f_j \tag{90}$$

[22]The use of the improved estimator of eq.78 in the box approximation is prohibited by the quadruple sums that it would contain.

This form has the advantage of including all modes up to an arbitrary m without much effort, with the overhead, of course, of being quadratic in N. As N grows beyond 10^5 this becomes particularly impractical. For investigating purposes, however, this approach is useful in testing the behavior of $E_2^{(q)}$ with more modes included (that is presumably the small λ limit).

It is remarkable that in the limit $m \to \infty$ we have $\psi(\vec{x}_i - \vec{x}_j) = M\delta_{i,j}$, and this leads to $s = 1$

$$E_2^{(q)} = \left(\frac{1}{N^2} \sum f_i^2 - \frac{1}{N^3} (\sum f_i)^2 \right) \tag{91}$$

In that limit a good point-set would have to integrates well any mode using a finite number of points N. Since that is impossible, all point-sets will be evaluated as equally bad by the particular diaphony.

It is evident that one has to find an optimal value for m. In the following plot the estimator $E_2^{(q)}$ is shown for TF5 in 2 dimensions with different values for m ranging from 3 to 30.

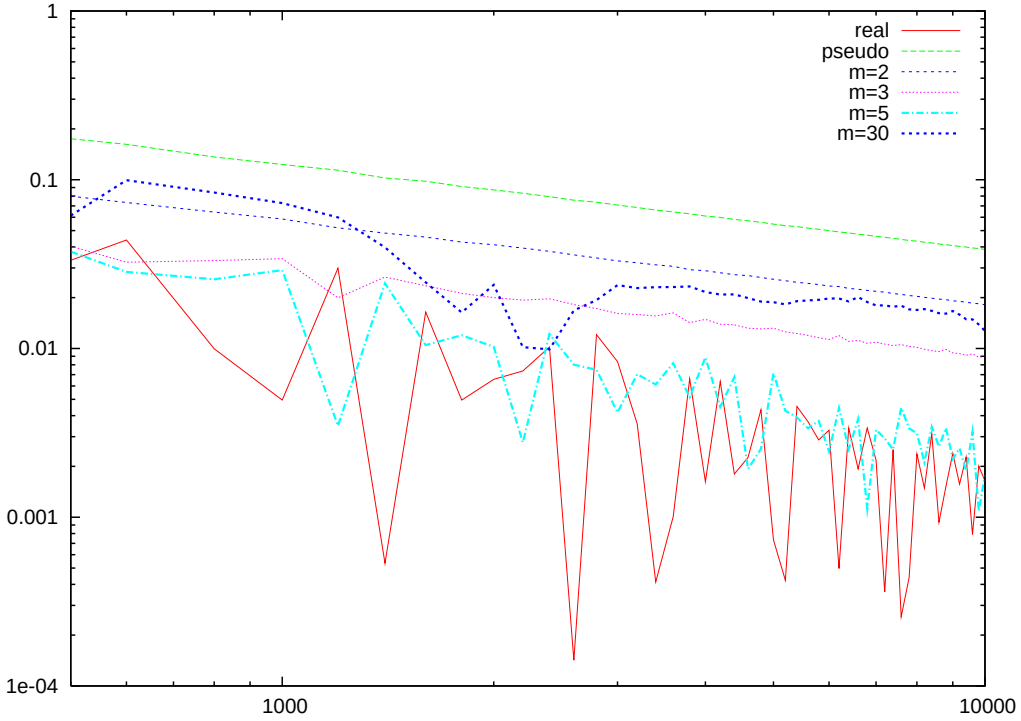


Figure 10: TF6, d=2 log-plot of the real error, and then from top to bottom the classical estimate, E_2^q with $m = 2$, with $m = 3$ and $m = 5$. The more modes one adds to the estimator the better it behaves. We also include the case $m = 30$ (orange line), to demonstrate that there is a turning point in m above which the estimate becomes worse. Note that $m = 5$ means square length up to $2m^2 = 50$, much higher than 15 that was our ceiling in the plots of the previous sections.

5 Concluding remarks

- The use of Quasi-Monte Carlo point-sets in numerical integration achieves a smaller error than the use of pseudo-random Monte Carlo point-sets. This advantage cannot be put in use without a reliable method for estimating the integration error.

- The 'classical', stochastic, error estimator relies on the assumption that the points in the point-set are uncorrelated. When used with a Quasi-Monte Carlo point-set, this assumption no longer holds. We saw that this leads to overestimating the error, thereby canceling any advantage gained by using the Quasi-Monte Carlo point-set.

- An estimator of stochastic nature is still possible but the underlying ensemble can not be the ensemble of all point-sets. We advocate the use of the ensemble of point-sets with the same degree of uniformity, as measured by a chosen diaphony. This approach leads to a prescription for a correlation function and an estimator, without the use of any information on the particular point-set or integrand.

- The price to pay is the raise in the computational complexity of the estimator from linear to quadratic in the number of points, which reflects the inclusion in the estimator of correlations between pairs of points. Using properties of diaphonies one can revert to a complexity that is linear times the number of modes involved.

- The error estimator suggested in this paper is shown to perform better than the 'classical' error estimator, resulting in an estimate up to an order of magnitude smaller than the 'classical' one.

- The flexibility of the construction (reflected in the freedom to choose the precise diaphony and the number of modes included) allows one to trade accuracy for computational cost. In computationally expensive applications, the monitoring approach of section 4.2 could be used to obtain an estimate that lies somewhere between the 'classical' and the quasi regime.

A number of further investigations have to be undertaken before implementing Quasi-Monte Carlo in the demanding field of phase space integration in particle physics. We defer these and further testing of the error estimator suggested above, in realistic cases, to further work.

acknowledgement

We would like to thank dr.C.Papadopoulos for persistently reminding us that a constant function can always be integrated with zero error.

References

[1] Douglas Adams, "The Hitchhiker's Guide to the Galaxy", Pan London, 1979.

[2] Hoogland and Kleiss Comp.Phys.Comm.98:128-136,1996 [hep-ph/9601270]

[3] Hoogland and Kleiss Comp.Phys.Comm.101:21-30,1997 [hep-ph/9609244]

[4] M.Luescher, Comp.Phys.Comm.,79,100 (1994)
F.James,Comp.Phys.Comm.,79,110 (1994)

[5] Halton, Num.Math. 2, 84-90 (1960)

[6] Ch.Schlier Comp.Phys.Comm., 159,93 (2004)

[7] H.Niederreiter J.Number Theory 30 (1988) 51-70

[8] P.Bratley, B.Fox and H.Niederreiter AMC Transactions on Modeling and Computer Simulation 2,No.3 (1992),195-213

[9] A.Owen, Monte Carlo extension of Quasi Monte Carlo, Winter Simulation Conference, IEEE Press, 1998

Appendix A: Estimators by diagrammatics

Diagrammatics for Quasi-Monte Carlo and Monte Carlo

Our strategy for obtaining the form of the estimators is best described by an example. Consider the triple sum

$$S_{p_1} S_{p_2} S_{p_3} \equiv \sum_{i,j,k=1}^{N} f_i^{p_1} f_j^{p_j} f_k^{p_3} \quad . \tag{92}$$

In our approach we need to compute the expectation value of this object including the first sub-leading order in $1/N$. It is given by

$$\begin{aligned}
\langle S_{p_1} S_{p_2} S_{p_3} \rangle &= N^{\underline{3}} \int f_i^{p_1} f_j^{p_j} f_k^{p_3} \left(1 - \frac{1}{N} \left(F_2(i,j) + F_2(i,k) + F_2(j,k) \right) \right) \\
&\quad + N^{\underline{2}} \int \left(f_i^{p_1+p_2} f_k^{p_3} + f_i^{p_1+p_3} f_j^{p_2} + f_i^{p_1} f_j^{p_2+p_3} \right) + \mathcal{O}(N) \\
&\approx N^3 \int f_i^{p_1} f_j^{p_j} f_k^{p_3} - N^2 \int f_i^{p_1} f_j^{p_j} f_k^{p_3} \left(\alpha_{ij} + \alpha_{ik} + \alpha_{jk} \right) \\
&\quad + N^2 \int \left(f_i^{p_1+p_2} f_k^{p_3} + f_i^{p_1+p_3} f_j^{p_2} + f_i^{p_1} f_j^{p_2+p_3} \right) \quad , \qquad (93)
\end{aligned}$$

with implied integration over the subscripts. The sub-leading terms in the expectation value are, therefore, obtained by either connecting any two of the summands in the multiple sum Ω with a factor $-\alpha$, or by contracting them. Now, any estimator E consists of a linear combination of terms like the above. Its variance, $\langle E^2 \rangle - \langle E \rangle^2$, contains both leading and sub-leading terms. The leading terms, however, cancel completely, and so do the sub-leading terms coming from a connection/contraction *inside* one of the factors E. We arrive at the following diagrammatic prescription. A sum of powers of f will be represented by a labeled dot, and a connection (including the $-\alpha$) by a link between dots. For example,

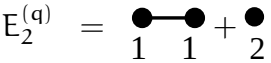
$$= \sum_{i,j,k,l=1}^{N} f_i^3 f_j f_k^4 f_l^2 \alpha_{jk} \alpha kl \quad . \qquad (94)$$

Now, suppose that the estimator E is given as a linear combination of *connected* diagrams. The estimator of its variance is the given by the *connected sub-leading* diagrams that can be obtained from $E \times E$. The factors $1/N$ can be added in a straightforward manner: each sum with p different summing indices carries a factor N^{-p}, and there is an additional overall factor N^{1-2^k} in E_{2^k}.

Estimators for Quasi-Monte Carlo

We apply the above considerations to the first estimators $E^{(q)}_{1,2,4}$ for Quasi-Monte Carlo. Squaring and constructing the connected sub-leading diagrams, we find

$E_1^{(q)} =$ 

$E_2^{(q)} =$ 

$$E_4^{(q)} = 4\,\underset{1\ 1\ 1\ 1}{\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet} + 4\,\underset{1\ 2\ 1}{\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet} + 4\,\underset{2\ 1\ 1}{\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet} + 4\,\underset{3\ 1}{\bullet\!\!-\!\!\bullet} + \underset{2\ 2}{\bullet\!\!-\!\!\bullet} + \underset{4}{\bullet} \quad . \tag{95}$$

Upon insertion of the correct factors of $1/N$, we arrive precisely at the estimators $E_{1,2,4}^{(q)}$ given in this paper. The construction of $E_8^{(q)}$ is straightforward: at that order, tree diagrams with branches develop. It may be worth noting that in this diagrammatic approach it becomes immediately clear that no diagrams with loops (that is, occurrences of α_{jj}, or $\alpha_{ij}\alpha_{ji}$, or $\alpha_{ij}\alpha_{jk}\alpha_{ki}$, and so on) are possible to this order in $1/N$.

Estimators for Monte Carlo

The MC estimators are of course precisely those of Quasi-Monte Carlo, with the replacement $\alpha_{ij} \to 1$. This means that the topology of the tree diagrams becomes irrelevant, and we can feasibly go up to E_{16}. We find

$$E_K = \frac{1}{N^{2K-1}} \sum_{s=0}^{K-1} E_{K,s} N^s \quad , \quad K = 1, 2, 4, 8, 16 \ , \tag{96}$$

where the coefficients of the various powers of N are given by

$$E_{1,0} = S_1 \ , \tag{97}$$
$$E_{2,0} = -S_1^2 \ ,$$
$$E_{2,1} = S_2 \ , \tag{98}$$
$$E_{4,0} = -4S_1^4 \ ,$$
$$E_{4,1} = 8S_1^2 S_2 \ ,$$
$$E_{4,2} = -S_2^2 - 4S_1 S_3 \ ,$$
$$E_{4,3} = S_4 \ , \tag{99}$$
$$E_{8,0} = -256S_1^8 \ ,$$
$$E_{8,1} = 1024S_1^6 S_2 \ ,$$
$$E_{8,2} = -1152S_1^4 S_2^2 - 512S_1^5 S_3 \ ,$$
$$E_{8,3} = 352S_1^2 S_2^3 + 832S_1^3 S_2 S_3 + 224S_1^4 S_4 \ ,$$
$$E_{8,4} = -4S_2^4 - 224S_1 S_2^2 S_3 - 128S_1^2 S_3^2 - 208S_1^2 S_2 S_4 - 96S_1^3 S_5 \ ,$$
$$E_{8,5} = 32S_2 S_3^2 + 8S_2^2 S_4 + 48S_1 S_3 S_4 + 48S_1 S_2 S_5 + 32S_1^2 S_6 \ ,$$
$$E_{8,6} = -S_4^2 - 8S_3 S_5 - 4S_2 S_6 - 8S_1 S_7 \ ,$$

$$
\begin{aligned}
E_{8,7} &= S_8 \ , \qquad (100)\\
E_{16,0} &= -4194304S_1^{16} \ ,\\
E_{16,1} &= 33554432S_1^{14}S_2 \ ,\\
E_{16,2} &= -104857600S_1^{12}S_2^2 - 16777216S_1^{13}S_3 \ ,\\
E_{16,3} &= 162922496S_1^{10}S_2^3 + 93585408S_1^{11}S_2S_3 + 7733248S_1^{12}S_4 \ ,\\
E_{16,4} &= -132579328S_1^8S_2^4 - 189530112S_1^9S_2^2S_3 - 20185088S_1^{10}S_3^2\\
&\quad -37552128S_1^{10}S_2S_4 - 3538944S_1^{11}S_5 \ ,\\
E_{16,5} &= 54444032S_1^6S_2^5 + 172064768S_1^7S_2^3S_3 + 69861376S_1^8S_2S_3^2 + 63553536S_1^8S_2^2S_4\\
&\quad +15532032S_1^9S_3S_4 + 14942208S_1^9S_2S_5 + 1507328S_1^{10}S_6 \ ,\\
E_{16,6} &= -9806848S_1^4S_2^6 - 69660672S_1^5S_2^4S_3 - 77729792S_1^6S_2^2S_3^2 - 45197312S_1^6S_2^3S_4\\
&\quad -43855872S_1^7S_2S_3S_4 - 5931008S_1^8S_3S_5 - 21135360S_1^7S_2^2S_5\\
&\quad -622592S_1^9S_7 - 5357568S_1^8S_2S_6 - 8060928S_1^7S_3^3 - 2802688S_1^8S_4^2 \ ,\\
E_{16,7} &= 551936S_1^2S_2^7 + 14180352S_1^5S_2S_3^3 + 10500096S_1^3S_2^5S_3 + 3200614 4S_1^4S_2^3S_3^2\\
&\quad +12816384S_1^4S_2^4S_4 + 6193152S_1^6S_2S_4^2 + 36679680S_1^5S_2^2S_3S_4\\
&\quad +7016448S_1^6S_3^2S_4 + 13725696S_1^6S_2S_3S_5 + 11722752S_1^5S_2^3S_5\\
&\quad +2007040S_1^7S_4S_5 + 6072320S_1^6S_2^2S_6 + 1994752S_1^7S_3S_6\\
&\quad +250880S_1^8S_8 + 1798144S_1^7S_2S_7 \ ,\\
E_{16,8} &= -256S_2^8 - 6438912S_1^3S_2^2S_3^3 - 3819520S_1^2S_2^4S_3^2 - 366592S_1S_2^6S_3\\
&\quad -1046016S_1^2S_2^5S_4 - 3568128S_1^4S_2^2S_4^2 - 8730624S_1^4S_2S_3^2S_4\\
&\quad -2233344S_1^3S_2^4S_5 - 1807360S_1^5S_3S_4^2 - 9879552S_1^3S_2^3S_3S_4\\
&\quad -8638464S_1^4S_2^2S_3S_5 - 2035712S_1^5S_3^2S_5 - 342016S_1^6S_5^2 - 2471936S_1^4S_2^3S_6\\
&\quad -3492864S_1^5S_2S_4S_5 - 1542144S_1^5S_2^2S_7 - 570368S_1^6S_2S_8 - 618496S_1^6S_3S_7\\
&\quad -607232S_1^6S_4S_6 - 851968S_1^4S_3^4 - 96256S_1^7S_9 - 3602432S_1^5S_2S_3S_6 \ ,\\
E_{16,9} &= 542208S_1S_2^4S_3S_4 + 2359296S_1^2S_2^2S_3^2S_4 + 1404160S_1^3S_2S_3S_4^2\\
&\quad +1608704S_1^3S_2^2S_3S_6 + 514432S_1^2S_2^3S_4^2 + 924672S_1^4S_3S_4S_5\\
&\quad +765952S_1^4S_2S_4S_6 + 552960S_1^2S_2S_3^4 + 1405952S_1^2S_2^3S_3S_5\\
&\quad +1814528S_1^3S_2S_3^2S_5 + 540672S_1S_2^3S_3^3 + 1394688S_1^3S_2^2S_4S_5\\
&\quad +262656S_1^2S_2^4S_6 + 808960S_1^4S_2S_3S_7 + 90624S_1S_2^5S_5 + 575488S_1^3S_3^3S_4\\
&\quad +451584S_1^4S_2S_5^2 + 191488S_1^5S_5S_6 + 475136S_1^4S_3^2S_6 + 164864S_1^5S_4S_7\\
&\quad +422912S_1^3S_2^3S_7 + 179200S_1^5S_3S_8 + 343296S_1^4S_2^2S_8 + 167936S_1^5S_2S_9\\
&\quad +1024S_2^6S_4 + 33792S_1^6S_{10} + 133760S_1^4S_4^3 + 60416S_2^5S_3^2 \ ,
\end{aligned}
$$

$$
\begin{aligned}
E_{16,10} &= -174336S_1S_2^2S_3S_4^2 - 191488S_1S_2S_3^3S_4 - 49920S_1^2S_2S_4^3 - 120832S_1S_2^3S_3S_6 \\
&\quad -290304S_1^2S_2S_3^2S_6 - 183936S_1^2S_2^2S_4S_6 - 242688S_1S_2^2S_3^2S_5 - 99328S_1S_2^3S_4S_5 \\
&\quad -87040S_1^3S_4^2S_5 - 231936S_1^2S_2^2S_3S_7 - 134400S_1^3S_2S_4S_7 - 153728S_1^3S_2S_3S_8 \\
&\quad -174080S_1^3S_2S_5S_6 - 64512S_2^3S_3^2S_4 - 105216S_1^2S_3^2S_4^2 - 172288S_1^3S_3S_4S_6 \\
&\quad -29184S_2^4S_3S_5 - 100352S_1^2S_3^3S_5 - 105472S_1^3S_3S_5^2 - 111360S_1^2S_2^2S_5^2 \\
&\quad -90112S_1^3S_3^2S_7 - 22528S_1S_2^4S_7 - 40000S_1^4S_4S_8 - 47104S_1^4S_5S_7 \\
&\quad -68608S_1^3S_2^2S_9 - 49088S_1^2S_2^3S_8 - 47616S_1^4S_3S_9 - 42240S_1^4S_2S_{10} \\
&\quad -10752S_1^5S_{11} - 1152S_2^4S_4^2 - 23552S_2^2S_3^4 - 512S_2^5S_6 - 23296S_1^4S_6^2 \\
&\quad -456960S_1^2S_2S_3S_4S_5 - 16384S_1S_3^5 \ , \\
E_{16,11} &= 3328S_2^3S_5^2 + 20352S_2^2S_3S_4S_5 + 16576S_1S_2S_4^2S_5 + 25088S_1S_3^2S_4S_5 \\
&\quad +27136S_1S_2S_3S_5^2 + 4096S_3^4S_4 + 33024S_1^2S_3S_5S_6 + 26240S_1^2S_2S_3S_9 \\
&\quad +32768S_1S_2S_3^2S_7 + 17536S_1S_2^2S_4S_7 + 5184S_1S_3S_4^3 + 24064S_1^2S_3S_4S_7 \\
&\quad +25856S_1^2S_2S_5S_7 + 19456S_1S_2^2S_3S_8 + 16448S_1^2S_2S_4S_8 + 20224S_1S_2^2S_5S_6 \\
&\quad +11392S_2S_3^2S_4^2 + 13440S_1^2S_2S_6^2 + 11328S_1^2S_4^2S_6 + 10240S_2S_3^3S_5 \\
&\quad +16000S_1^2S_4S_5^2 + 224S_2^4S_8 + 11520S_2^2S_3^2S_6 + 832S_2^3S_4S_6 + 13312S_1S_3^3S_6 \\
&\quad +13568S_1^2S_3^2S_8 + 9472S_1^3S_6S_7 + 6912S_2^3S_3S_7 + 9984S_1^3S_5S_8 \\
&\quad +8832S_1^3S_2S_{11} + 10368S_1^3S_3S_{10} + 4224S_1S_2^3S_9 + 8576S_1^3S_4S_9 \\
&\quad +10176S_1^2S_2^2S_{10} + 352S_2^2S_4^3 + 46336S_1S_2S_3S_4S_6 + 3008S_1^4S_{12} \ , \\
E_{16,12} &= -768S_3^2S_5^2 - 2944S_2S_3S_5S_6 - 1024S_3S_4^2S_5 - 2208S_1S_2S_5S_8 - 1664S_1S_2S_4S_9 \\
&\quad -1216S_2S_4S_5^2 - 3072S_1S_2S_3S_{10} - 2944S_2S_3S_4S_7 - 3584S_1S_3S_5S_7 \\
&\quad -2944S_1S_2S_6S_7 - 2016S_1S_3S_4S_8 - 3008S_1S_4S_5S_6 - 768S_1^2S_7^2 - 1024S_3^3S_7 \\
&\quad -1536S_3^2S_4S_6 - 1472S_1^2S_6S_8 - 224S_2S_4^2S_6 - 1920S_1S_3S_6^2 - 1024S_1S_4^2S_7 \\
&\quad -1472S_2S_3^2S_8 - 1408S_2^2S_5S_7 - 208S_2^2S_4S_8 - 960S_1S_2^2S_{11} - 1792S_1^2S_3S_{11} \\
&\quad -1440S_1^2S_2S_{12} - 1312S_1^2S_4S_{10} - 1792S_1S_3^2S_9 - 1088S_2^2S_3S_9 \\
&\quad -1856S_1^2S_5S_9 - 768S_1S_5^3 - 128S_2^2S_6^2 - 704S_1^3S_{13} - 4S_4^4 - 96S_2^3S_{10} \ , \\
E_{16,13} &= 128S_2S_7^2 + 128S_5^2S_6 + 32S_4S_6^2 + 160S_3S_5S_8 + 48S_2S_6S_8 + 8S_4^2S_8 \\
&\quad +256S_3S_6S_7 + 192S_4S_5S_7 + 128S_3^2S_{10} + 224S_1S_5S_{10} + 160S_2S_5S_9 \\
&\quad +128S_3S_4S_9 + 192S_1S_6S_9 + 160S_1S_7S_8 + 224S_1S_3S_{12} + 48S_2S_4S_{10} \\
&\quad +32S_2^2S_{12} + 192S_2S_3S_{11} + 128S_1S_4S_{11} + 160S_1S_2S_{13} + 128S_1^2S_{14} \ , \\
E_{16,14} &= -S_8^2 - 16S_5S_{11} - 16S_7S_9 - 8S_6S_{10} \\
&\quad -4S_4S_{12} - 16S_3S_{13} - 16S_1S_{15} - 8S_2S_{14} \ ,
\end{aligned}
$$

$$E_{16,15} \;=\; S_{16}\ . \tag{101}$$

The number of individual terms in each E_K is that of the partitions $\Pi(K)$ of K: $\Pi(1) = 1$, $\Pi(2) = 2$, $\Pi(4) = 5$, $\Pi(8) = 22$, and $\Pi(16) = 231$. Likewise, the number of terms in each $E_{K,s}$ is the partition of K into $(K - s)$ parts. We have not extended our results to the fifth-order error estimator with $K = 32$ and $\Pi(32) = 8349$, since already E_8 and E_{16} are purely academic and we have included them only as an illustration of the method.

Appendix B: The $O(\frac{1}{N^2})$ contribution to G_p

The second order contribution to G_p can be found by summing up $O(\frac{1}{N^2})$ terms coming from

1. the pure rings (containing only 2-point vertices)
2. the three graphs contributing to $G_p^{(1,2,3)}$ (containing one 4-vertex, two 3-vertices or two external points)
3. products of a pure ring and one of the three graphs above or two of the graphs above.
4. the new graphs (containing one 6-vertex,one 5-vertex and one 3-vertex, two 4-vertices, one 4-vertex and two 3-vertices, four 3-vertices, one 3-vertex and three external points, two 3-vertices and two external points or one 4-vertex and two external points)

After a lengthy but straightforward calculation (involving some cancellations) we get

$$\begin{aligned}
G_p^{(3)} \;=\; & \tfrac{G_0 p}{N^2} \;\Big(\; \frac{2p-1}{4}K_3 - \frac{1}{4}(K_2)^2 - \frac{1}{2}K_2(x_i,x_j) \;\Big) \\
& + \tfrac{G_0}{N^2} \;\Big(\; \frac{3}{8}K_5 + \frac{1}{3}K_4 - \frac{1}{32}K_3^2 - \frac{1}{8}K_2K_1(x_i,x_j) \\
& \quad - \frac{1}{2}K_3(x_i,x_j) - \frac{1}{12}L_{1,1,1} - \frac{1}{2}L_{2,1,1} - \frac{1}{4}L_{2,2,1} \\
& \quad - \frac{1}{48}K_3L_{1,1,1} - \frac{1}{4}L_{3,1,1} + \frac{1}{8}K_1(x_i,x_j)^2 + \frac{1}{2}Q_1(x_i,x_j,x_k) \\
& \quad + \frac{1}{4}Q_2(x_i,x_j) + \frac{1}{2}Q_3(x_i,x_j,x_k) + \frac{1}{4}Q_4(x_i,x_j)
\end{aligned}$$

$$+\frac{1}{6}L_{1,1,1}(x_i,x_j,x_k)+\frac{1}{24}L_{1,1,1}K_1(x_i,x_j)+\frac{1}{288}L_{1,1,1}^2$$
$$+\frac{1}{48}M_1+\frac{1}{8}M_2+\frac{1}{24}M_3+\frac{1}{16}M_4)$$

where

$$K_{a,b,\ldots}\equiv\sum_{1,2,\ldots}\rho_1^a\rho_2^b\ldots \tag{102}$$

$$K_a(x_i,x_j)\equiv\sum_{i,j}{}'\sum_{1}\rho_1^a e_{\vec{n}_1}(x_i)e^*_{\vec{n}}(x_j) \tag{103}$$

$$L_{a,b,c}\equiv\sum_{1,2,3}\rho_1^a\rho_2^b\rho_3^c\delta_{1+2+3} \tag{104}$$

$$Q_1(x_i,x_j,x_k)\equiv\sum_{i,j,k}{}'\sum_{1,2}\rho_1\rho_2 e_{\vec{n}_1}(x_i)e^*_{\vec{n}_1}(x_j)e_{\vec{n}_2}(x_j)e^*_{\vec{n}_2}(x_k) \tag{105}$$

$$Q_2(x_i,x_j)\equiv\sum_{i,j}{}'\sum_{1,2}\rho_1\rho_2 e_{\vec{n}_1}(x_i)e^*_{\vec{n}_1}(x_j)e^*_{\vec{n}_2}(x_i)e_{\vec{n}_2}(x_i) \tag{106}$$

$$Q_3(x_i,x_j,x_k)\equiv\sum_{i,j,k}{}^{i}\sum_{1,2,3}\rho_1\rho_2 e_{\vec{n}_1}(x_i)e^*_{\vec{n}_1}(x_j)e_{\vec{n}_2}(x_j)e^*_{\vec{n}_2}(x_k)e_{\vec{n}_3}(x_j)e^*_{\vec{n}_3}(x_k) \tag{107}$$

$$Q_4(x_i,x_j)\equiv\sum_{i,j}\sum_{1,2}\rho_1^2\rho_{1+2}e_{\vec{n}_1}(x_i)e^*_{\vec{n}_1}(x_j) \tag{108}$$

$$L_{a,b,c}(x_i,x_j,x_k)\equiv\sum_{i,j,k}{}'\sum_{1,2,3}\rho_1^a\rho_2^b\rho_3^c e_{\vec{n}_1}(\vec{x}_i)e_{\vec{n}_2}(\vec{x}_j)e_{\vec{n}_3}(\vec{x}_k)\delta_{1+2+3} \tag{109}$$

$$Q_{a,b}(x_i,x_j,x_k)\equiv\sum_{i,j,k}{}'\sum_{1,2}\rho_1^a\rho_2^b e_{\vec{n}_1}(\vec{x}_i)e^*_{\vec{n}_1}(\vec{x}_j)e_{\vec{n}_2}(\vec{x}_j)e^*_{\vec{n}_2}(\vec{x}_k) \tag{110}$$

$$M_1\equiv\sum_{1,2,3,4}\rho_1\rho_2\rho_3\rho_4\delta_{1+2+3+4} \tag{111}$$

$$M_2\equiv\sum_{1,2,3,4,5}\rho_1\rho_2\rho_3\rho_4\rho_5\delta_{1+2-5}\delta_{3+4-1-2}\delta_{5-3-4} \tag{112}$$

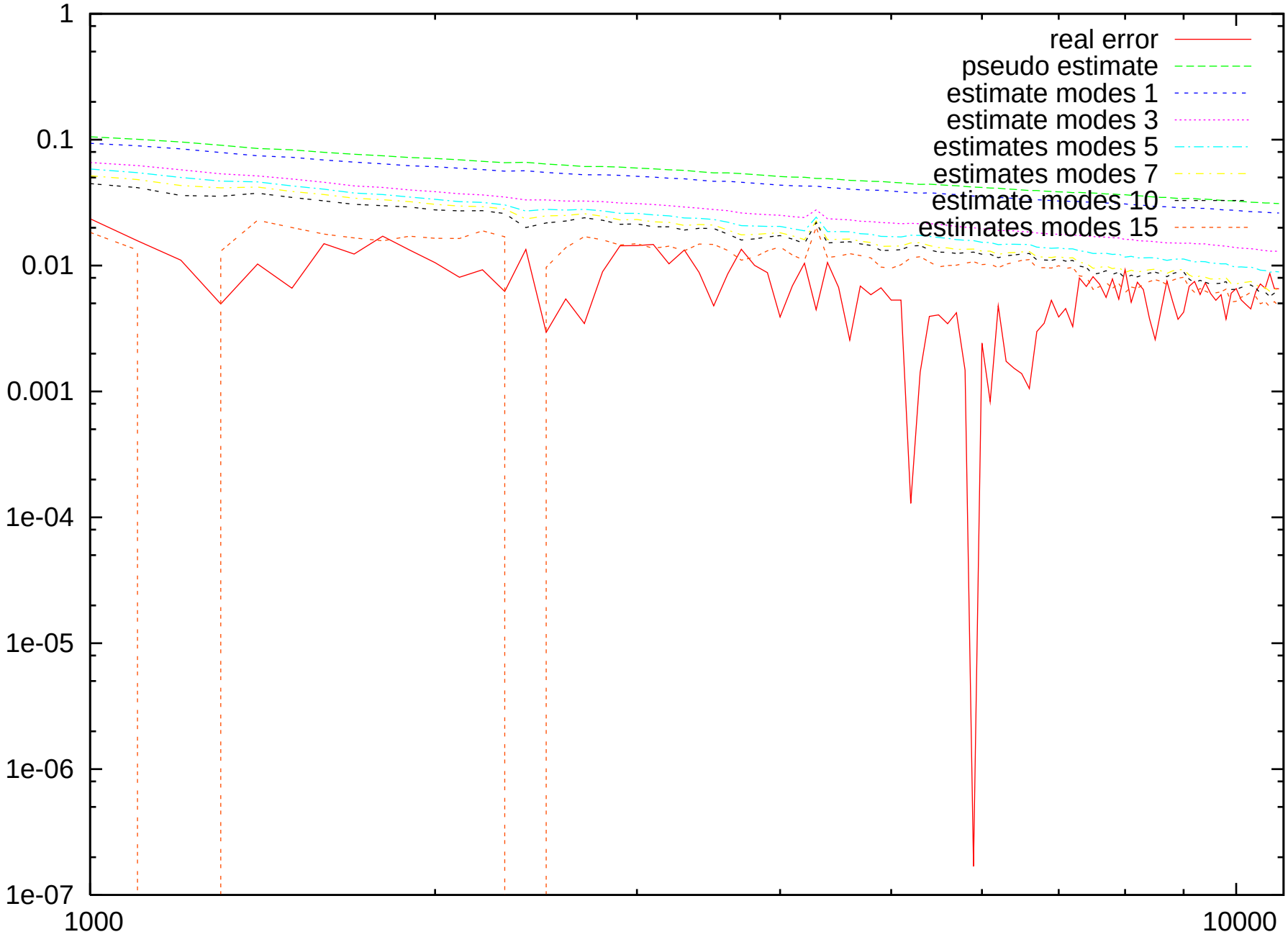
$$M_3\equiv\sum_{1,2,3,4,5,6}\rho_1\rho_2\rho_3\rho_4\rho_5\rho_6\delta_{1-2-5}\delta_{2-3-6}\delta_{3-1-4}\delta_{4+5+6} \tag{113}$$

$$M_4 \equiv \sum_{1,2,3,4,5,6} \rho_1\rho_2\rho_3\rho_4\rho_5\rho_6\delta_{1-2-5}\delta_{2-3-6}\delta_{3+6-4}\delta_{4-5-1} \tag{114}$$

with

$$\rho_i \equiv \frac{2z\sigma^2_{\vec{n}_i}}{1-2z\sigma^2_{\vec{n}_i}} \tag{115}$$

and $\sum^i_{i,j,k,\ldots} = \sum_{\vec{x}_i \neq \vec{x}_j \neq \vec{x}_k \ldots}$, $\sum_{1,2,\ldots} \equiv \sum_{\vec{n}_1,\vec{n}_2,\ldots}$ and $\delta_{1+8-2+\ldots} \equiv \delta(\vec{n}_1 + \vec{n}_8 - \vec{n}_2 + \ldots)$.


1
0.1
0.01
0.001
1e-04
1e-05
1e-06
1e-07
1000
10000
real error
pseudo estimate
estimate modes 1
estimate modes 3
estimates modes 5
estimates modes 7
estimate modes 10
estimates modes 15